Effects of Feedback Templates on Student Performance

Ryley Russell

MacEwan University

Abstract

We provided one of three types of feedback to students who performed poorly on short written assignments and provided an opportunity to revise and resubmit their work. Control feedback pointed out the main problems of the submission, ‘strong template’ feedback added an annotated example of a good assignment, and ‘weak template’ feedback added an annotated example of a poor assignment. “Strong” feedback led to a greater likelihood of a passing grade on resubmitted work, compared to weak or control feedback templates, but only when we accounted for student motivation. In fact, if a “less motivated” student received the strong template, they were 2.3 times more likely to be successful in their resubmission, but “more motivated” students were likely to be successful no matter what feedback they received.

Keywords: templates, student performance, feedback

Effects of Feedback Templates on Student Performance

Feedback is an integral part of the educational process. Without feedback, students are unable to discern which aspects of their work require repair or improvement (Boud & Molloy, 2013). Feedback is utilized as an educational tool to allow students to obtain information about their performance. Moreover, it allows them to recognize the parallels and differences between the works they have produced and that which is considered ideal (Boud & Molloy, 2013). These benefits are important for both the aforementioned reasons and their meta-analytically supported significance (Black & Wiliam, 1998).

Though feedback is widely regarded as an integral part of the educational process, the mechanism by which it is delivered and which form of feedback is optimal has received little attention. In fact, in a recent literature review of *PsychInfo*, a search for the word "feedback" returned 39,000 hits; however, only 36 of these articles were found with the words "written feedback" in the title (Jolly & Boud, 2010). These findings illustrate an important gap within current literature as students frequently express dissatisfaction with the varying quality of feedback and the mechanisms by which it is delivered (Hattie & Timperley, 2007). In order to optimize the usefulness of feedback, it is important to understand how it functions in an academic context and the different perspectives by which it can be considered.

Feedback as an educational tool can be considered from a social learning perspective or from operant conditioning theory. Operant conditioning is a type of learning in which the consequences of a given behaviour determine the future probability of its occurrence. Distilled, this theory holds that if a behaviour is rewarded its future probability increases and if it is punished then the future probability decreases (Powell, Honey, & Symbaluk, 2013). Much of the early research surrounding this theory can be traced back to B.F. Skinner who operated within

the behaviourism school of psychology. Behaviourists assert that research should examine the environmental antecedents and consequences that lead to and explain an observable behaviour (Skinner, 1988). For example, if we were scared we would not hide because we are fearful; rather, we would hide because some event has occurred that caused us to hide and simultaneously feel fearful. In this example, the event may be aggressive posturing or a loud noise that is acting as the catalyst for both the observable behavioural response and the associated feelings. In essence, the environmental antecedent (aggressive posturing or loud noise) simultaneously explains why we behave in such a way and why we feel a certain way (Skinner, 1988). The implications of this theory have been extended to the educational settings in the classroom; however, these are usually enacted with the aim of increasing or decreasing the likelihood of certain classroom behaviours (Altman & Linton, 1971).

From the operant conditioning perspective, feedback is the consequence of academic behaviour. Rather than influence overt classroom behaviours with rewards like verbal praise or reprisal, feedback allows the student to understand which responses are satisfactory and which are not. This contributes to a phenomenon known as shaping. That is, the feedback gradually leads to a higher quality response through the reinforcement of successive approximations of the ideal response (Powell et al., 2013). In addition to this behavioural approach is the cognitive-behavioural viewpoint of social learning theory. This theory emphasizes observational learning and the cognitive variables that can assist in explaining human behaviour (Powell et al., 2013). Within this theory are the principles of reciprocal determinism and modeling; both of which can assist in elucidating the potential benefits of correctly given feedback.

The principle of reciprocal determinism holds that overt behaviour, cognitive/affective variables, and environmental events all influence each other (Janousek, 1992). For example, let’s

assume a student raises their hand in class to answer a question posed by the instructor. The student attempts an answer but provides an incorrect response and is laughed at by classmates. This laughter in turn results in feelings of embarrassment within the responding student. The environmental event (question posed by the instructor) leads to an overt behaviour (answer attempt) and due to a second environmental event (class laughter) the student experiences a negative cognitive/affective response (embarrassment). Let us also assume that after the classroom ceases its laugher, the instructor provides feedback and ultimately corrects the initial response.

This combination of an affectively loaded scenario coupled with corrective feedback make the student significantly more likely to remember the correct response in the future (Groome, 2014). This is in part due to the heightened attention that comes with being affectively stimulated (Baddeley, Eysenck, & Anderson, 2015; Ohman, Flykt, & Esteves, 2001). Another interesting factor is that it is not just negative affectivity which drives attention; the student derives attentional benefits if positive emotions are felt as well. In fact, the presence of positive affectivity broadens attentional resources and enhances flexible cognitive strategies (Fredrickson & Branigan, 2005). While this is more obviously a consideration for verbal praise and/or verbal feedback, it holds when receiving feedback in written form as well. The student who sees that they have received a bad grade may feel embarrassed or ashamed while the student who sees they have earned an exemplary grade may feel proud. We can see that when the student is affectively stimulated by the grade they have received, they are cognitively primed to pay more attention when feedback is given in written format. In doing so, the student is more likely to identify and incorporate important details from the feedback into their future responses. These

factors and benefits are important, but there is yet another piece of social learning theory that contributes to optimal student outcomes.

Modeling is a component of social learning theory that was initially observed in the imitation of aggressive behaviours through observational learning (Bandura, Ross, & Ross, 1961). In an academic context feedback provides a model upon which the student can rebuild their subsequent response. This is crucial, as modeling and shaping together form a dual effect to provide reinforcement for an improved response and provide the framework for an optimal future response. Each of these perspectives supports evidence to the assertion that feedback in written format can positively influence student performance. However, the quality of feedback will interact with student motivation to influence the extent to which student performance improves. For example, if the student is extrinsically motivated by nothing more than achieving a good grade, the feedback becomes nothing more than a means to an end. However, if the student is intrinsically motivated and interacts with the feedback conscientiously, they are more likely to connect with and understand the information it contains (Higgins, Hartley, & Skelton, 2002). In this latter scenario, the feedback becomes an end in itself. That is, the feedback contributes to the more intrinsically motivated goal of developing a deep understanding of the material. Although this is the case, the manner in which feedback is delivered influences its efficacy.

Feedback can be delivered via many different mechanisms and its form is often dependent upon context. In an educational context, feedback is most commonly given in written format (Jolly & Boud, 2010). The vast majority of teacher – student communications occur in this form and this represents a form of one – way communication. Not only is this type of communication unidirectional, but it also often lacks immediacy and invariably carries a high probability of being misunderstood (Jolly & Boud, 2010). Additional problems with

unidirectional feedback include the fact that the student is rarely able to follow up on it directly. For example, if a student has just received corrective feedback regarding a classical conditioning procedure, the best time for the student to employ the feedback is as soon after the initial attempt was made as possible and within a similar context. In fact, this opportunity for future practice coupled with a successive retrieval attempt leads to faster knowledge acquisition that can be transferred to different contexts (Roediger & Butler, 2011; Foss & Pirozzolo, 2017). However, it is worth noting that transferring a concept to a different context cannot occur if mastery of that concept never occurred. In this sense, keeping the context consistent while learning is still taking place is beneficial for the student; hitting a stationary target is easier than hitting one that is in motion.

If the aforementioned student struggling with the concept of classical conditioning isn't able to put corrective information into practice in the same context, they are unlikely to be able to do so successfully in the future or in a different context. Unfortunately, students are unlikely to face a similar question again until they're in an exam setting. This is to the detriment of the student, as not providing students with the opportunity to revisit material they haven't mastered does not facilitate academic skill acquisition or deep retention of the provided information (Groome, 2014). For example, let's assume a student is revisiting a question centred on material they haven't yet mastered, but with feedback present. The student is more likely to be able to recall the requisite information and answer the question correctly; even if the feedback is not given immediately. This is because the feedback is cognitively linked to the concept in question. This effect becomes more potent when the student is conscientious, intrinsically motivated to understand core concepts, and actively attempts to retrieve pertinent information (Groome, 2014;

Higgins et al., 2002). Though feedback has many positive benefits when used correctly, it can be harmful to student outcomes if used incorrectly (Hattie & Timperley, 2007).

Feedback is delivered to assist the student in understanding where they deviated from assignment expectations and provide a suggestion for how they could resolve the problem in the future. Unfortunately, feedback can leave students dissatisfied for a number of reasons. The primary complaint that students have concerning feedback is that it lacks direct specificity for ways in which they can improve (Higgens, Hartley, & Skelton, 2001). Moreover, it is imperative that the feedback be clear, understandable, and carries the potential to be immediately acted upon by students (Gibbs & Simpson, 2004). Constructive or corrective feedback should also be impersonal and professional. Corrective communication with students should be free of any ad hominem statements and be constructive rather than punitive. Feedback that is viewed as personal rather than constructive criticism by the students can result in damage to their self-perception and/or their self-confidence (Carless, 2006). Moreover, this perceived attack serves to remove the student's focus from the task in question (Kluger & DeNisi, 1996). The overarching goal of post-secondary institutions is to assist students in developing long-term retention of the material they are studying. Without focus or attention, there can be no encoding or retrieval. For example, if one friend tells another to remember the first three digits of a phone number, but the second friend wasn't paying attention to the directive, they will not be able to recall the desired information because it was never encoded (Groome, 2014). For these reasons, punitive or ad hominem based feedback must be avoided; it disturbs the student's attention. The student may remember what they perceive as an "attack" more than the constructive content of the feedback.

Another important consideration is the language used when delivering feedback. Though feedback commonly occurs as a unidirectional construct in academia, this is indicative of the

importance of accessible language within the feedback. In fact, Higgens and colleagues assert that students may often be unable to interpret, understand, and employ the suggestions found within feedback from their instructors (2001). If the student is unable to grasp a concept and demonstrates this by failing to perform adequately on a test or an assignment, unclear feedback only serves to exacerbate the student's issues. For example, if feedback is overly simplistic (ex. "you're wrong) it provides no indication for how the student could improve their response. In behaviourist terms, this provides a form of punishment, but no modeling. Moreover, if the feedback contains overly academic language that the student doesn't understand, the student is similarly unable to improve their response. Another factor relevant to overall performance and application of feedback concerns the amount of times a student attempts to actively retrieve pertinent information. By providing clear feedback and allowing multiple attempts at a question that tests the same core concept, the student becomes more likely to benefit from the testing effect.

The testing effect occurs when an individual repeatedly attempts to actively retrieve a memory and those repeated attempts subsequently facilitate successful retrieval (Groome, 2014). In fact, this act of repeated active retrieval has been shown to enhance long-term memory for the material over simply repetitively reading material (Baddeley et al., 2015). The effects of this phenomenon are enhanced the more it is utilized. In other words, the more an individual is tested, the more likely they are to successfully retrieve the target information (Bangert-Drowns, Kulik, & Kulik, 1991; McDaniel, Anderson, Derbish, & Morrisette, 2007). These are tangible and potent benefits to be sure, and additionally it is important to note that this effect applies across a wide array of different learning activities (Butler & Roediger, 2007; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). These various learning and retrieval activities include

recall, recognition, and enhanced organizational processes (Groome, 2014; Baddeley et al., 2015; Zaromb & Roediger, 2010). Research more relevant to the university context has shown that the testing effect can also have a positive influence when students attempt to recall concepts tested in different forms (Foss & Pirozzolo, 2017). One would expect that the testing effect would exert the strongest influence when a student is tested on the material in the same way it was presented; however, this was not always the case. Foss and Pirozollo (2017) found that a student was also likely to display enhanced retrieval capabilities if the item was seen in a short answer or multiple choice format first and then presented in the other format during subsequent testing. Moreover, the students were also more likely to retrieve information tested in previous quizzes in a re-worded format during a higher stakes test. These findings indicate that the testing effect not only works to enhance basic retrieval but can also enhance the ability of the student to apply previously learned knowledge to new problems (Foss & Pirozzolo, 2017).

As previously cited literature suggests, student complaints regarding feedback cluster around feedback lacking specificity, not possessing the potential to be immediately acted upon, and lacking clarity (Gibbs & Simpson, 2004; Higgens et al., 2001). The aims of the present study were twofold. First, we sought to address the complaints students frequently express about feedback by creating standardized feedback templates. Second, we sought to test the feedback templates to see if they were an effective tool to improve the quality of student written work and enhance academic outcomes. Moreover, ideal feedback should also be impersonal and professional (Carless, 2006). Each of these issues was covered by our creation of feedback templates.

Our templates were designed to be specific and clear so that students could understand where they made mistakes and how to fix them. The assignments were designed so that students

could revise and resubmit their work. This allowed them to immediately act on feedback. In addition, the language we used was intended to be professional and encouraging rather than personal and denigrating. Moreover, the feedback templates provided impersonal examples as models for resubmissions. A "strong" template provided one possible example of a good answer with graphics that highlighted important features. A "weak" template provided an example of a poor answer with graphics that highlighted the key missing or incorrect components of the answer. For example, if a student received the strong feedback template, the graphics that identify strong components of the answer provide a model for the student to base their resubmission. This provides the student with a type of response scaffold and contributes to the shaping of an ideal response. We hypothesized that this would ultimately contribute to improved written work and academic outcomes.

After using the templates, we examined the likelihood that a student would receive full marks on their resubmission as a function of the template they had received. At the outset of the study, we predicted that the strong response condition would lead to the greatest likelihood of a successful resubmission and the control condition would lead to the lowest likelihood of a successful resubmission. We predicted that the weak response condition would fall somewhere between the strong response condition and the control condition regarding its efficacy in assisting the students in improving on their original submission. We hypothesized that if a student receives a feedback template of any kind, they would have a greater likelihood of a successful resubmission relative to the control condition feedback. However, we also hypothesized that if a student received the strong response condition, then they would demonstrate a greater likelihood for a successful resubmission relative to the weak response condition. In addition, we also hypothesized that the templates would show effects that

generalized to subsequent reflection attempts. For example, if a student had received the strong feedback template for their second reflection, then we predicted they would be more likely to pass their third reflection on the first attempt. Much like the effects on resubmission attempts, we thought the strong template would have the strongest effect, the control feedback the weakest, and the weak template somewhere in between. This is to say that if a student received control feedback on their second reflection (and therefore had failed their first attempt), they would be the most likely to have to resubmit on their third reflection compared to a student who had received either the strong or weak feedback template.

Methods

Participants

Participants for this study were students enrolled in introductory psychology classes at MacEwan University. Participation in the study involved completing a no-stakes, optional reflection assignment following each weekly activity. There were 16 sections comprised of 120 seats for a highest possible sample of 1,920. At the time we gathered data, we had a sample of n=1,774. Of these students, 1,372 submitted at least one of the optional reflection assignments and 634 resubmitted at least one of the optional reflection assignments. Though the breakdown of male to female students for the psychology sections specifically is unknown, it likely closely mirrors the entire student body at the university. The student body at MacEwan is 63.2% female and 36.8% male with an average age of 22.7 years (“Facts and Figures,” n.d.).

Materials

The materials used in this study consisted of questions contained within the optional weekly reflection assignments and the associated templates (see appendix A for examples of

reflection questions). For each weekly activity, three types of feedback were created: a control condition, a weak condition, and a strong condition (see appendix B). These conditions were counterbalanced across the sixteen sections on a weekly basis. Counterbalancing essentially means the templates were distributed such that each section was guaranteed to receive the same number of each of the conditions.

Procedure

The students were given the option to complete a weekly reflection assignment following their in-class activity. The course was organized such that for every reflection a student completed, 1% would be guaranteed toward their final grade thereby reducing the weight of the final exam. This meant that participation was beneficial but was not necessary to achieve 100% in the course. For example, there were eight possible reflections that could be completed over the duration of the course. If all were completed, then 8% would be guaranteed toward the final grade. Consequently, the final exam would only be worth 32% of the final grade rather than 40%. Moreover, the reflections were phrased in the same manner as commonly encountered mid-term or final exam short answer questions. This allowed the students to connect the material from each activity to an application based question. This activity was performed in an open book setting and was therefore likely to lead to deeper retention and superior understanding (Nevid, Pyun, & Cheney, 2016). Moreover, the students were able to benefit from the testing effect after the activity, a practice that has been well-documented for its educational efficacy (Boud, 2001; Foss & Pirozzolo, 2017; Groome, 2014).

These reflections were graded on a 1-5 scale. It is important to note that these grades were assigned based on the subjective evaluation of a teaching assistant (TA) based on a rubric provided by the course creator. If the student scored four or more, the reflection was deemed

satisfactory and no revision was required. However, if the student received a score of three or lower out of five, they were assigned to one of the three conditions.

The control condition feedback advised the student to go back over their notes and the associated chapter(s), lecture videos, or visit the learning centre to refresh themselves on the material. The second condition represented a strong reflection. This reflection was illustrated with arrows and comments that allowed the student to see exactly why the answer in the template would be worth full marks. This gave them clues as to what they could incorporate in their updated reflection to elevate their mark. The third condition represented what would be considered a weaker reflection; that is, a reflection that would be worth a score of three or lower. This reflection was also illustrated with arrows and comments that allowed the student to see exactly why the template would be considered weak. The primary aspect of the weak template that differed from the strong template was that the strong template transparently presented and identified the components of one possible correct answer to the student. The weak template gave the students clues as to what they should look for to be included in a strong answer, but the actual components were not supplied via the template. One important note about the grading of the reflection resubmissions is that regardless of the condition, if the student again received a score of three or lower, no mark was awarded and their original score of 0 was secured; however, all students who received an initial score of 0 were told they could resubmit for full marks. Another measure that allowed the control template to be fairly compared against the strong and weak template conditions was the fact that all conditions provided the same standard feedback (to visit the learning centre and re-visit the textbook). The only difference was that a template was attached in the strong or weak feedback condition for the student to use while revising the failed reflection.

An important aspect of the templates is that they were administered by the teaching assistants; the same people responsible for viewing and marking the reflections. This provided a safeguard against plagiarism of the templates themselves. If a student was caught plagiarising a template, they would incur consequences related to academic integrity. After the students who received an initial score of three or lower resubmitted their reflections, we examined the data for two different outcomes. These were the likelihood of a successful resubmission depending on feedback template condition and the likelihood of successful submissions on future reflections.

Results

Although our original sample consisted of 1773 students in Introductory Psychology, not all students are included in our analysis of the effect of the feedback templates. See Table 1 for a summary of completion rates for reflections across four weeks of topics. Most submitted reflections passed on the first attempt (70.6%). Of the reflections that did not pass, and received feedback, 634 (45.8%) were revised and resubmitted before the deadline. It is those 634 resubmitted reflections that constitute our sample for the following analyses. Within the sample, 240 received the control condition (37.9%), 203 received the 'strong' feedback template (32%) and 191 received the 'weak' template (30%).

We performed logistic regression to predict the likelihood of a successful resubmission. The first model included the feedback condition (control/strong/weak), the student's attendance record for the four weeks of class (range of 1-4), the number of reflections completed during the four weeks of topics (range of 1-4), and which topic was covered in that reflection (coded as 1-4). Note that we used attendance and number of reflections completed as proxy measures of motivation or perhaps conscientiousness. The model was a good fit, revealing significant contributions of feedback condition ($p < .001$), number of reflections completed ($p < .001$), and

topic ($p < .001$), but not attendance ($p < .259$). Pairwise comparisons revealed that the strong template predicted greater success than did the weak or control template ($p < .023$) but no difference between the weak and control templates ($p < .152$). See Figure 1 for a representation of those success differences. Resubmitted reflections were successful 84.7% of the time after the strong template was used, whereas the weak template was associated with a 78% success and the control condition had 75% success upon resubmission. Most resubmitted reflections were successful, but those that followed the strong template were somewhat more likely to be successful.

We repeated the logistic regression analysis without attendance as a factor, and the remaining factors were again significant predictors of successful resubmission (all $p \leq .001$). Further, the strong template was again significantly different from the weak template ($p < .031$) but the weak template did not differ from control.

When we added the interaction term between reflections completed and feedback condition, the significant effect of the feedback condition was eliminated, with no significant interaction between the terms. See Figure 2 for a visual representation of the effect of feedback condition when different numbers of reflections have been completed. While the pattern is consistent across the four groups (strong > weak/control), the differences are large when a student completed few reflections and virtually non-existent when a student completed most or all reflections. For example, when all four reflections were completed by a student then almost all resubmitted reflections were successful, regardless of feedback condition (98.2% - 100%). When only one reflection was completed by a student, then the difference in likelihood of success between conditions ranged from 22.2% (in the control condition) to 65.5% (in the strong template condition). The effect of the template appears to be significant only when ‘motivation’

is accounted for, but the effect is not an interaction effect. Lastly, we also conducted logistic regression to determine whether any of the feedback conditions influenced the likelihood of a successful submission on the next immediate reflection; however, no condition was a significant predictor (all $p > .05$).

Discussion

At the beginning of this study, we predicted that the strong feedback template would carry the greatest likelihood of a successful resubmission relative to the weak or control feedback templates. We predicted the control template would carry the smallest effect, and the weak feedback template would fall somewhere between the strong and control condition templates. Each of these predictions proved correct following statistical analysis. The descriptive statistics in Figure. 1 show that in order of potency, the strong, weak, and control templates all increased the likelihood of a successful resubmission. In addition, the strong feedback template was significantly different from either the weak or control template in its ability to increase the likelihood of a successful resubmission. Interestingly, this was most pronounced in students who we classify as “less motivated” based on the number of reflections they attempted. With reference to our initial hypothesis that the templates would influence the likelihood for a successful first attempt on the next reflection, our analysis showed that none of the templates had a significant effect.

It is important to note that out of our total sample of 1,774; only 1,372 ever attempted the optional reflection assignments during the first six weeks of class and only 634 ever resubmitted. This means that out of our sample, there may have been a slight restriction of range problem with respect to student motivation. This is because there were still 359 students who never attempted a reflection. This indicates that although the feedback templates were most effective for the

students classified as less motivated, these students were still more motivated than the 359 students who never submitted anything. This means we may have observed an even more pronounced effect of the feedback templates if these even less motivated students had attempted the reflection assignments and subsequently resubmitted.

In this study, we also aimed to ameliorate the main complaints students have about feedback. These included unnecessary variance in feedback quality and receiving feedback that could be misunderstood or lacks specificity. In addition, we wanted to determine which feedback template possessed the best content to help students. By using templates, we were able to ensure that each student was receiving standardized feedback. With this standardization comes a sharp decrease in variance. Additionally, even though we used three distinct types of feedback, the results of this study indicate that if instructors opt to use feedback templates they are best served by using strong feedback templates. If this is done, the instructor has effectively eliminated most variance in feedback. This creates a level playing field for the students, increases instructor efficiency, and allows for useful feedback even within larger classes.

The second common student complaint regarding clarity of feedback was addressed by incorporating explanations within the templates themselves. These explanations allow the student to understand precisely what made a response strong or weak (depending on template) and most importantly, how to utilize that recommendation in their subsequent response. However, because the students were getting the opportunity to re-attempt the same assignment, we can’t be sure if the high successful resubmission rates are due to the testing effect, a ceiling effect, the feedback templates, a student’s motivation, or any combination of these.

Admittedly, this study was inherently noisy. This is to say that because there are several factors that affect student performance, developing an experiment to ascertain the effects of

feedback templates is problematic and will almost invariably lead to results that are questionable in their precision. For example, out of our feedback conditions, the control condition saw the greatest amount of use. This was not by design, but a by-product of confusion. During data screening the reflections were individually examined to ensure no late submissions were counted in the final sample and to ensure the correct feedback template was attached for each section and reflection. Unfortunately, in many instances this did not occur. The reflections were graded by several teaching assistants (TAs); each of whom had a unique grading style and different levels of motivation. Unfortunately, human error happens and errors were made that affected the quality of the feedback that was separate from that contained within the templates. In other cases, no feedback template was attached or the wrong template was attached. If the former scenario occurred, that datum was categorized within the control condition.

For example, in the control condition the TAs were supposed to identify any errors made and recommend that the student re-visit the relevant chapters, lecture videos, and to go to the learning centre to receive additional help from an instructor or TA. During data screening it was discovered that in some instances this protocol was not followed. Instead, the TA may have left a large amount of highly detailed feedback. In fact, in many instances the feedback contained direct advice on what should be included in the resubmission to ensure full marks were received. This practice is normally ideal for the student; however, in this case it may have undermined any effect (or lack thereof) that the control condition may have had. Essentially, because the control condition was variable, we may have observed a smaller effect of the templates than we would have if the control condition had been a true control. Unfortunately, at this point we can only speculate about the true differences in effects between conditions.

Another problem with this study was the fact that there was no way to confirm the student had looked at the template. During data screening, we could see exactly how the student's successful resubmission was different from their initial submission. It appeared that some students did not use or did not understand the feedback because their resubmission offered no substantial improvement. Thus, some students in template conditions may not have actually received the manipulation. For these reasons, out results are not entirely clear.

A third issue involves the classification of students as either passing or failing their initial submissions or their resubmissions due to variance in grading by individual TAs. While each reflection possessed an associated grading rubric, there was never a point where the TAs got together to grade sample submissions to ensure adequate inter-rater reliability. Ideally this meeting to grade submissions would have occurred each week so all the TAs were clear on how they were to grade the reflections for each topic. However, because this never occurred we can only speculate about the inter-rater reliability. Again, the size of our sample and the random assignment of TAs to grade certain sections likely provides a safeguard from any major influence. However, this is another example of a lack of control in our study that could obscure the results.

One final issue with the methodology was the fact that the TAs responsible for grading the reflections were not blind to the condition that the students were in. Because of unintentional bias, it is possible that any number of the TAs either graded too leniently or too harshly on the original submission or the resubmission depending on their expectancies. Admittedly, the likelihood of this occurring is low, but it is a complication nonetheless. If we were to replicate or extend this study, the inter-rater reliability concerning the grading of the reflections is undoubtedly more important than blindness concerning the grader(s). However, any issues

relevant to the methodology that contribute to the overall robustness of the results are worth mentioning and rectifying.

This study involved a relatively understudied topic in educational psychology and future research will ideally strive to resolve the errors that were present to increase the reliability and validity of the results. These flaws were not mentioned to imply that imperfect methodology is permissible; rather, in the hopes that an identification of these errors would increase the reliability of future research in the same area. By working towards improved methodology, we can increase the robustness of our findings and see the results put into practice. Ideally this will assist in optimizing student learning outcomes. Although the data set had many influencing factors, these all dilute the effects of the templates and decrease the likelihood of finding significant results. For these reasons, it is likely that if the methodological errors were rectified, an even greater effect of the templates would be observed.

The results of this study illuminate several implications or applications for the feedback templates. For example, future research could determine if the use of feedback templates generalizes to other assignments or exams. Even though we did not observe a transfer of any positive effects to subsequent reflections or different topics, it is possible that there would be improvements on similar topics/exam questions related to the original topic. Additionally, research could examine how effective the templates are in remedial settings, as student motivation and achievement are closely correlated (Goodman et al., 2011). To combine the correlation between student motivation and achievement with the results of the present study, an interesting question for future research is if the effect of the templates helps a weak student versus an unmotivated student the most. As previously mentioned, the best predictor of academic performance is previous academic performance. Unfortunately, because our sample consisted of

predominately first year university students we had no basis for a fair comparison to past academic performance. Though motivation and academic performance are correlated, they are not a measure of the same construct. For this reason, future research in this area could investigate which of these populations would receive greater benefits from the use of strong feedback templates.

References

Altman, K. I., & Linton, T. E. (1971). Operant conditioning in the classroom setting: A Review of the Research. *The Journal of Educational Research, 64,* 277-286.

Baddely, A., Eysenck, M. W., & Anderson, M. C. (2015). *Memory*. New York, NY: Psychology Press

Bandura, A., Ross, D., & Ross, S. (1961). Transmission of aggression through imitation of aggressive models. *Journal of Abnormal and Social Psychology, 63,* 575 – 582.

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. L. (1991). Effects of frequent classroom testing. *Journal of Educational Research, 61,* 213-238.

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice, 5,* 7-74.

Boud, D. (2001). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22,* 151-167.

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The Challenge of Design. *Assessment & Evaluation in Higher Education, 38,* 698-712.

Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 5*, 514-527.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education, 31,* 219-233.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14,* 4-58.

"Facts and Figures." MacEwan University (n.d.). Retreived from: https://www.macewan.ca/wcm/Discover/OurStory/FactsandFigures/index.htm

Foss, D. J., & Pirozzolo, J. W. (2017, March 23). Four Semesters Investigating Frequency of Testing, the Testing Effect, and Transfer of Training. *Journal of Educational Psychology*. Advance online publication. http://dx.doi.org/10.1037/edu0000197

Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition & Emotion, 19*, 313-332.

Gibbs, G. & Simpson, C. (2004) Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education, 1*, 3–31.

Goodman, S., Jaffer, T., Keresztesi, M., Mamdani, F., Mokgatle, D., Musariri, M., Pires, J., & Schlechter, A. (2011). An investigation of the relationship between students' motivation and academic performance as mediated by effort. *South African Journal of Psychology, 41*, 373-385.

Groome, D. (2014). *An Introduction to Cognitive Psychology: Processes and Disorders* (3rd ed.). New York, NY: Psychology Press

Hattie, J., and Timperley, H. (2007). The Power of feedback. Review of Educational Research, 77, 81-112.

Higgins, R., Hartley, P. & Skelton, A. (2001). Getting the message across: the problem of communicating assessment feedback. *Teaching in Higher Education, 6*, 269–274.

Higgins, R., Hartley, P. & Skelton, A. (2002). The conscientious consumer: Reconsidering the Role of Assessment Feedback in Student Learning. *Studies in Higher Education, 27*, 53-64.

Janousek, J. (1992). Bandura, Albert social cognitive theory. *Ceskoslovenska Psychologie, 36*, 385-398.

Jolly, B., & Boud, D. (2010). *Feedback in Higher and Professional Education.* New York, NY: Routledge

Kluger, A. N., & DiNisi, A. (1996). The effects of feedback interventions on performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Psychological Bulletin, 119,* 254-284.

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 5,* 494-513.

Nevid, J. S., Pyun, Y. S., & Cheney, B. (2016). Retention of text material under cued and uncued recall and open and closed book conditions. *International Journal for the Scholarship of Teaching and Learning, 10,* 1-4.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology, 130,* 466-478.

Powell, R. A., Honey, L. P., & Symbaluk, D. G. (2013). *Introduction to Learning and Behaviour* (4th ed.). Belmont, CA: Wadsworth

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20-27.

Skinner, B. F. (1988). The operant side of behaviour therapy. *Journal of Behavior Therapy and Experimental Psychiatry, 19,* 171-179.

Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition, 38,* 995-1008.

Table 1

Reflection Completion Rates

Activity:	Number Submitted	Number Failed	Number Resubmitted
Scientific Thinking	1367	468	215
Research Methods	1043	291	110
Neurotransmission	1142	322	161
Perception	1150	302	148
Totals:	4702	1383	634

Appendix A

Sample reflection questions.

Reflection #1 Question

Take a few minutes to think about beliefs that you have about studying. Do you have particular things that you do all the time because you think they're effective? Read over the section in your textbook about hypotheses and about falsifiability. Now create a hypothesis about one of your beliefs. Your answer should be in the form of a specific, falsifiable prediction about your belief and it should provide a clear indication of how you would measure your results. For this reflection, describe your beliefs in a couple of sentences, and then state your hypothesis in one clear sentence. (It might be helpful to include a sentence that briefly describes how you would test your hypothesis, if that isn't clear from your hypothesis itself.)

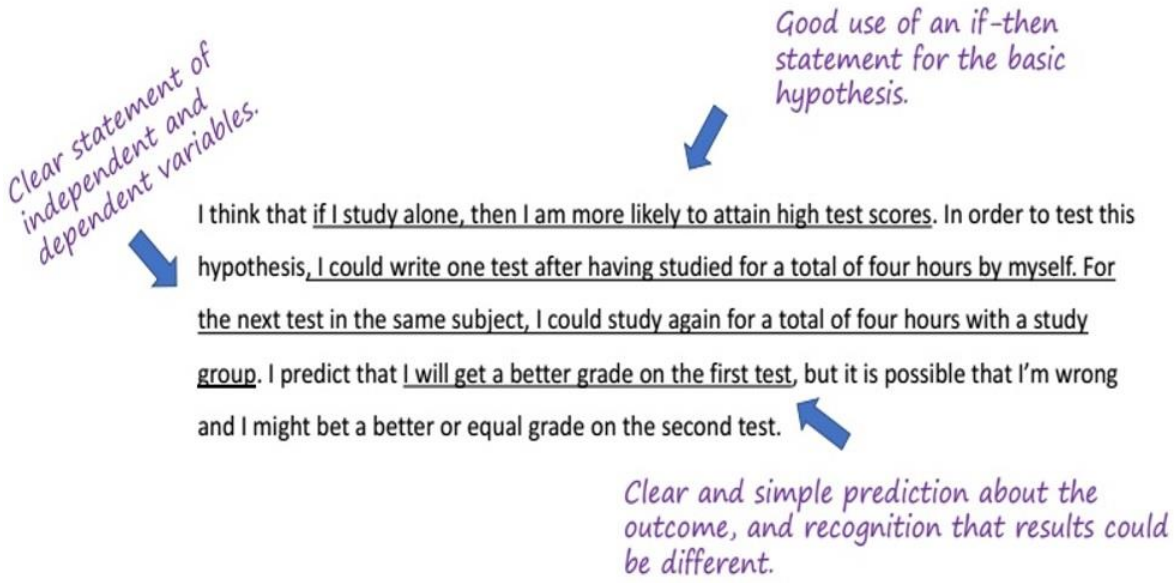
Reflection #3 Question

Epinephrine (also known as adrenaline) acts as a neurotransmitter but also acts as a hormone. How would epinephrine work differently as a hormone compared to how it works as a neurotransmitter? (See pages 105-107 in the textbook to help you answer this question). You should be able to answer this question in a sentence or two.

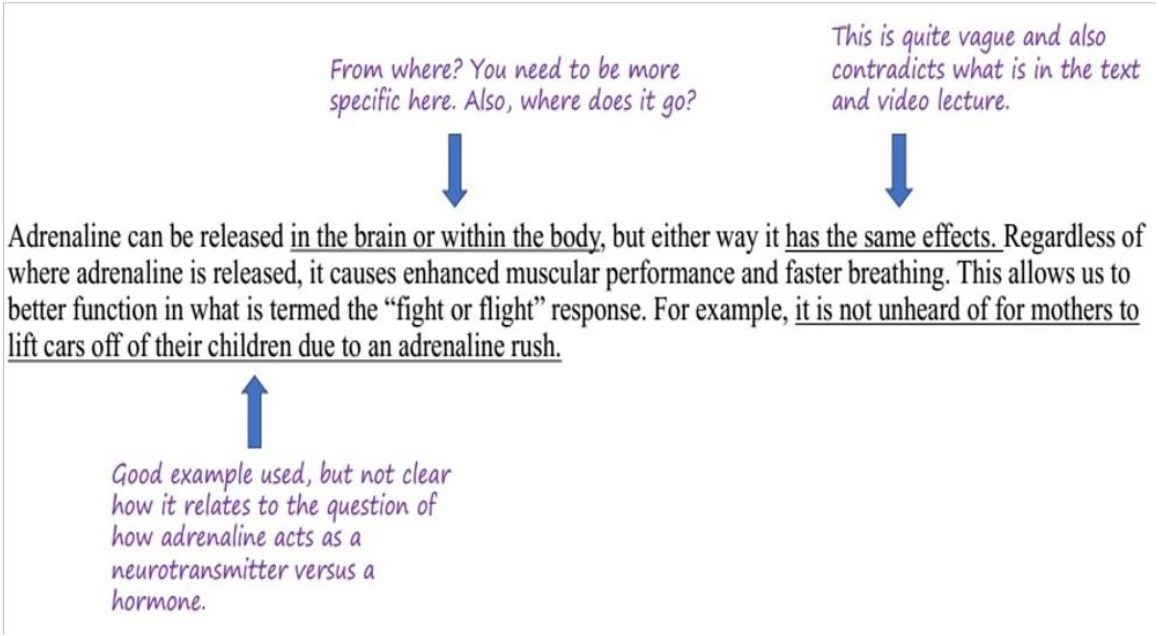
Appendix B

Sample strong and weak feedback templates.

Reflection #1 strong feedback template



Reflection #3 weak feedback template



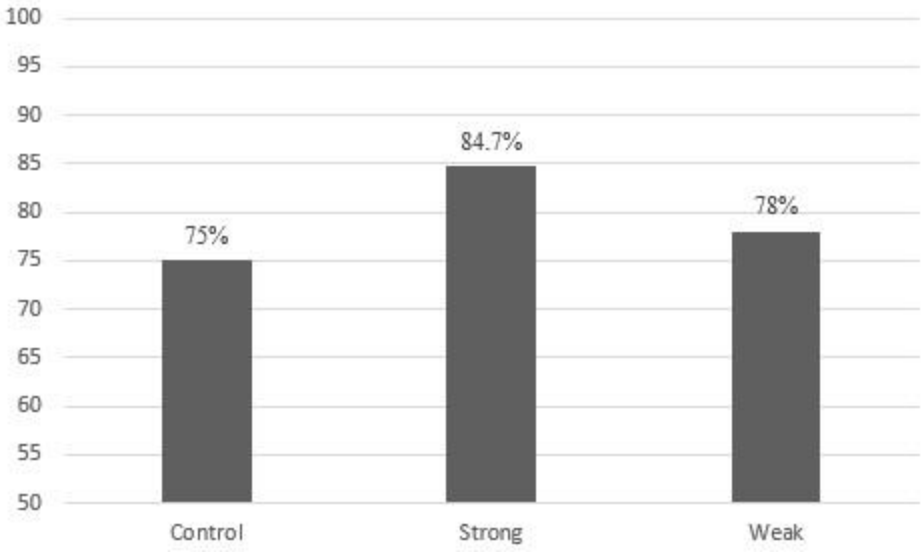


Figure 1. Likelihood of a successful resubmission across all four reflections depending on feedback template.

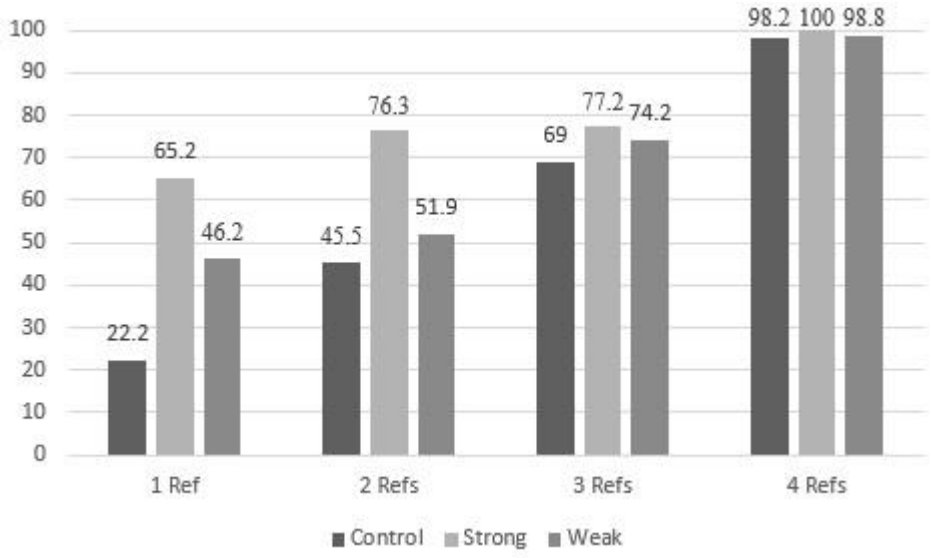


Figure 2. The effect (in percentages) of the feedback template influencing the likelihood of a successful resubmission dependent on number of reflections completed. Ex. 3 Refs = 3 reflection assignments completed prior to midterm.