Advanced Topics In Personalized Online Learning
05-499
Pooja Casula & Anita Delahay
Project Final Report

Introduction

Feedback and its implications in learning have been one of the most discussed topics in *Personalized Online Learning* 2. The general consensus among the feedback research studied is that, for the most part, feedback is effective and can lead to significant learning gains. In order to be effective, feedback should be immediate, positive, relate to the correct responses, and help students understand the goals of their practice (1,2).

However, there remains debate on which *form* of feedback is the most effective, across different domains and populations, and in different environments. Many studies have supported the general idea that more holistic feedback is more effective in terms of learning gains,

One 1990 study done by Mckendree looked at three forms of feedback to compare their effectiveness. The study compared minimal feedback, which simply informed the participant whether their answer was correct or incorrect, condition-violation feedback, which targeted and corrected the specific error made by the participant, and goal feedback, which informed the participant about what their final problem-state should be. Mckendree found that goal feedback was better than condition violation feedback, implying that making students aware about the objective of the problem is more helpful than simply correcting the specific error they made, and supporting the idea that providing more holistic feedback is generally more favorable (3).

Another study done by Sleeman et al. found that just reteaching the material when students made an error was as effective as correcting errors plus reteaching. The results of this study imply that correcting student errors may not have any significant impact on how well they understand the error, similar to the results found by Mckendree. The researchers in this study concluded that the reteaching aspect of the feedback was more helpful to students, again supporting the general trend that holistic feedback is better (4).

However, both of the studies described used human tutors, which differs from the environment we are interested in, namely intelligent tutoring systems. A 2015 meta-analysis done by Van der Kleij et al., exclusive to studies done in computer-based environments, found that providing elaborated feedback to students when they answer incorrectly produced an effect size of 0.49, which was significantly higher than the effect sizes produced from simple correctness feedback (0.05) and providing the correct answer (0.32). Again, elaborated feedback may be considered a form of holistic feedback and seems to be most effective (5).

While analyzing the previous research on contrasting different forms of feedback in the field, we found a few inconsistencies. We found that studies differed in terms of whether feedback is delivered via computer or human tutor (3,4,5). The method of administration may or may not impact how students process the feedback. We also found that the feedback conditions are not fully isolated from each other. For example, in the Sleeman et al. study, there was not a condition for only correct-error feedback but rather reteaching was included in all conditions, making it difficult to truly contrast the effectiveness of the feedback conditions. Finally, in the Van der Kleij et al. meta-analysis, their analysis did not include studies that gave error-correction feedback, which is clearly relevant in intelligent tutoring system environments. Identifying such inconsistencies in these previous studies led us to formulate our own research question.

For our final project, we decided to investigate feedback types in self-directed environments in which no instructor is present. This type of environment is typical for personalized and adaptive learning. In order to do this, we experimentally contrasted three conditions that varied only in the type of feedback delivered. We chose the conditions to have high contrast, relevance to online tutoring systems, and to be easily describable to other practitioners in the field. We also decided to vary only one variable, the feedback, and use a novel domain (i.e., experimental design in the field of psychology) for our experiment.

To contrast the effectiveness of the forms of feedback, we created three tutors in CTAT, identical in form and appearance, but different in the type of feedback returned.

Each tutor returned only one type of feedback, and our experimental design was between subjects. Therefore, subjects received only one type of feedback across all the questions they answered. We implemented three questions in each tutor module, each with four sub-questions (*Figure 1*). The sub-questions should be considered as separate questions that reference a common scenario, rather than all part of a single question (i.e., these are not “inner loop” questions, which allow for multiple paths through a problem). Finally, In order to limit confounding variables that could impact how students interpreted the feedback administered to them, we did not include hints in our tutor design, which may be considered another form of feedback to students (1).

The feedback types were:

1. **Condition 1: Simple correctness:** “That is correct.” or “That is incorrect.” (see *Figure 2*)
2. **Condition 2: Error-specific:** “His usual powder is one level of the independent variable, but not the independent variable.” (see *Figure 3*)
3. **Condition 3: Elaborated/Teaching**: “The independent variable causes the effect on the dependent variable. An independent variable must have at least two levels so that you can manipulate whether a subject gets one level or the other, compare effects.” (see *Figure 4*)

Based on past studies, we hypothesize that elaborated feedback will lead to the greatest learning gains, followed by error-specific feedback, followed by simple feedback.

The experiment was administered to nine Carnegie Mellon Students with varying degrees of knowledge regarding psychology and experimental design. We first administered a written Pre-Test to the participants (See Pre-Test/Post-Test in Final Deliverables). Next, students were logged in to the tutor that would run with their randomly assigned feedback condition. Finally, we administered a written post-test which was exactly the same as the pre-test. The purpose of administering a pre and post test was to detect a change in knowledge state, normalized by how much knowledge could be gained (Post - Pre/ Points Possible - Pre).

Results

Due to the low number of subjects (n = 9), with three subjects assigned per condition, we will report group means, standard deviations, and overall trends, rather than inferential statistics. The pre- and post-tests were scored against a rubric created before the experiment. Only one of the paper authors scored the tests, and so no inter-rater reliability measures are reported. Beyond this pilot, we will have had two raters and included these metrics. We first looked at simple Pre- to Post-Test gains.

The Pre- to Post-Test Gain data show a trend of the highest gain for the Elaborated condition, followed by Error Correction, then Simple Feedback, as had been hypothesized. However, the standard deviation of scores for the Elaborated feedback condition is the largest. This is due in part to the small *n*, but may also signal that this result is highly variable. See the Table below and *Figures 5 and 6* in the Appendix for graphical representations of these data.

Group	Mean Gain (Std. Dev.)
Simple Feedback	M = 0.34 (0.76)
Error Correction	M = 0.67 (0.58)
Elaborated / Teaching	M = 1.83 (1.04)

Another way to look at these data are by normalizing Pre- to Post-Test gains by the number of points the subject could gain by the instruction. This measure is more sensitive to the gains that can be made by more advanced learners, namely those who are able to score higher on the Pre-Test and thereafter close the gap between that score and the Post-Test. These data also show the trend of the highest gain for the Elaborated condition, followed by Error Correction, then Simple Feedback. In addition, the standard deviation of scores for the Elaborated feedback condition is this time the smallest, which supports the conclusion that this was indeed the best performing condition. See the Table and *Figures 7 and 8* in the Appendix for graphical representations of these data.

Group	Mean Gain (Std. Dev.)

Simple Feedback	M = 0.12 (0.44)
Error Correction	M = 0.39 (0.35)
Elaborated / Teaching	M = 0.47 (0.21)

Discussion

Work on cognitive tutors suggests that feedback should be short and error-directed (1), but feedback that is holistic or elaborated in nature (e.g., teaching of goal structure, general principles related to the material) has been empirically more advantageous when compared with condition violation feedback (3) or error-correction feedback (4) under other (i.e. human tutor) conditions. While this work did not contrast intelligent and human tutors directly, it did contrast three forms of likely feedback (i.e., teaching, error-directed, and simple feedback) within an intelligent tutor, and early results indicate that the most effective type of feedback when using human tutors (elaborated) is also most effective in the intelligent tutoring environment for this population (i.e., CMU students) and at this level of instruction. (i.e., introductory-level materials).

Contributions

This study contributes to the existing literature by contrasting three types of feedback which are a natural fit within the format of an intelligent tutor, easily implementable by instructors, instructional designers, and other "lay" users of intelligent tutoring systems.

In addition, we implemented the tutor in a novel area (i.e. psychology) and on a novel topic (i.e., experimental design).

Limitations

We recognize that there are limitations to both our tutor design and our experimental design. While creating our tutors, we came across two challenges. The first had to do with possible problem types that we were able to implement in the tutor. For example, in the original plan, we were to implement a problem in which students were asked to select all the possible right choices. We were unable to implement such a problem since the error-specific condition required each possible wrong answer choice to receive its own feedback message and CTAT required that we write an error message for every single permutation of the wrong answer. A second challenge was that we could not access the tutor log data generated by TutorShop, the site hosting the tutors. Since we were unable to access the log data,, we could not evaluate learner behavior while working in the tutor. Specifically, we could not contrast behaviors across conditions based on greater or less learning gains, and could not evaluate how the feedback condition may have influenced behaviors (i.e., whether more clicks per problem were recorded in certain conditions).

We also identified limitations based on the nature of our experiment. First, we had a very small sample size, so making inferences from our results is not possible. Second, because we did not have an inner loop to our questions, our recommendations, at best, would apply to personalizing feedback at the problem-level. Third, we did not equate the length of feedback, and therefore the overall time of exposure to feedback. Thus, participants in the Elaborated feedback condition likely received more instructional time. In the future, using tutor log data will allow us to control for the length of instructional exposure. The fourth limitation was that although the literature surveyed suggested five possible forms of feedback, we did not include two of them, namely feedback which tells the correct answer or feedback which explains the structure of the problem. Fifth, results may shift over time, so a delayed test condition would help determine any feedback effects on retention and whether they differ from the immediate test condition. The final limitation we identified was that we did not test for mastery level by feedback interactions. This would help us understand if certain feedback is best for students at different stages of learning.

Conclusion

As hypothesized, elaborated feedback was the most advantageous form of feedback in our small pilot study. If results of a larger study continued along current trends, we would be most interested in determining whether any competency level by feedback type interactions were present. In particular, we wonder whether more novice learners would benefit most from more elaborated feedback, while more advanced learners would benefit from error-specific or simple feedback. If that were the case, we may recommend Elaborated Feedback until a certain level of mastery is reached, and then transition to shorter forms of feedback, which would be more efficient (time-wise) as well as serve to eliminates misconceptions and refine existing knowledge structures.

Appendix

Figure 1. Four problems implemented in one question of one tutor module

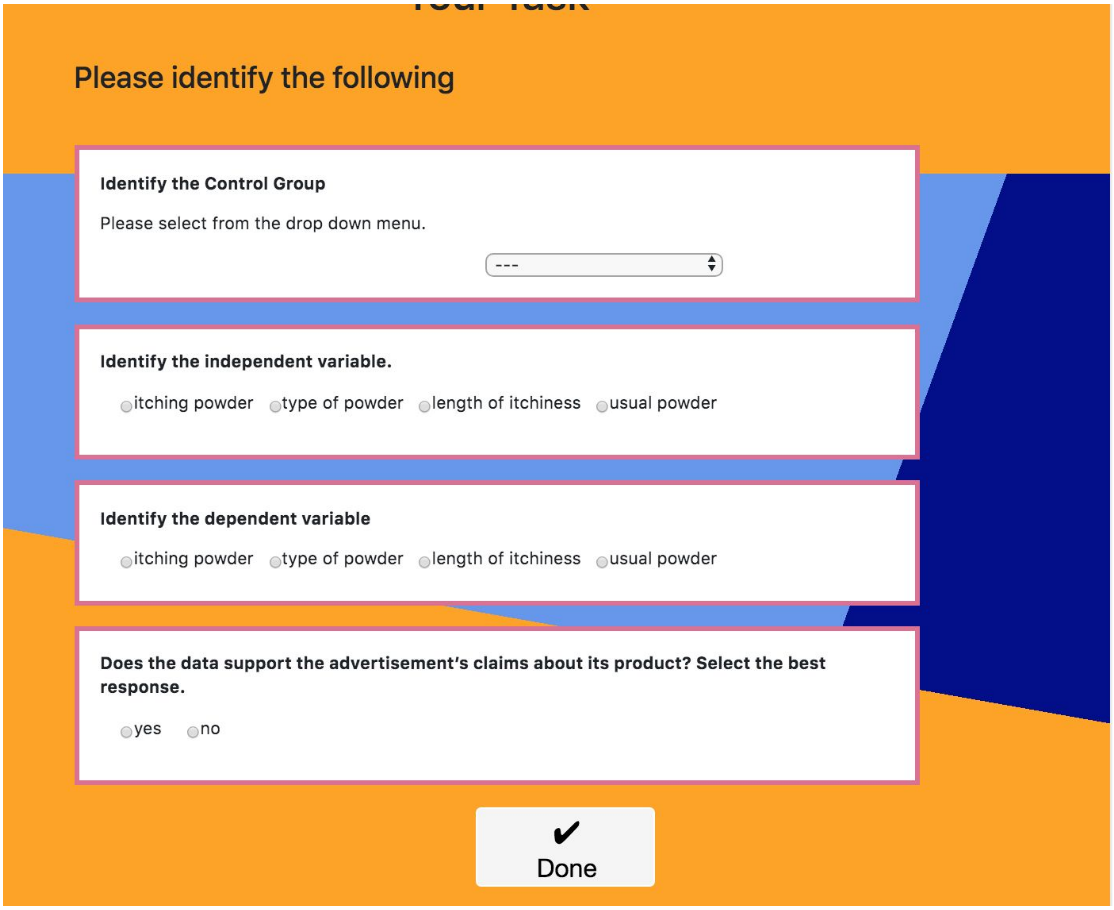


Figure 2. Feedback Condition 1

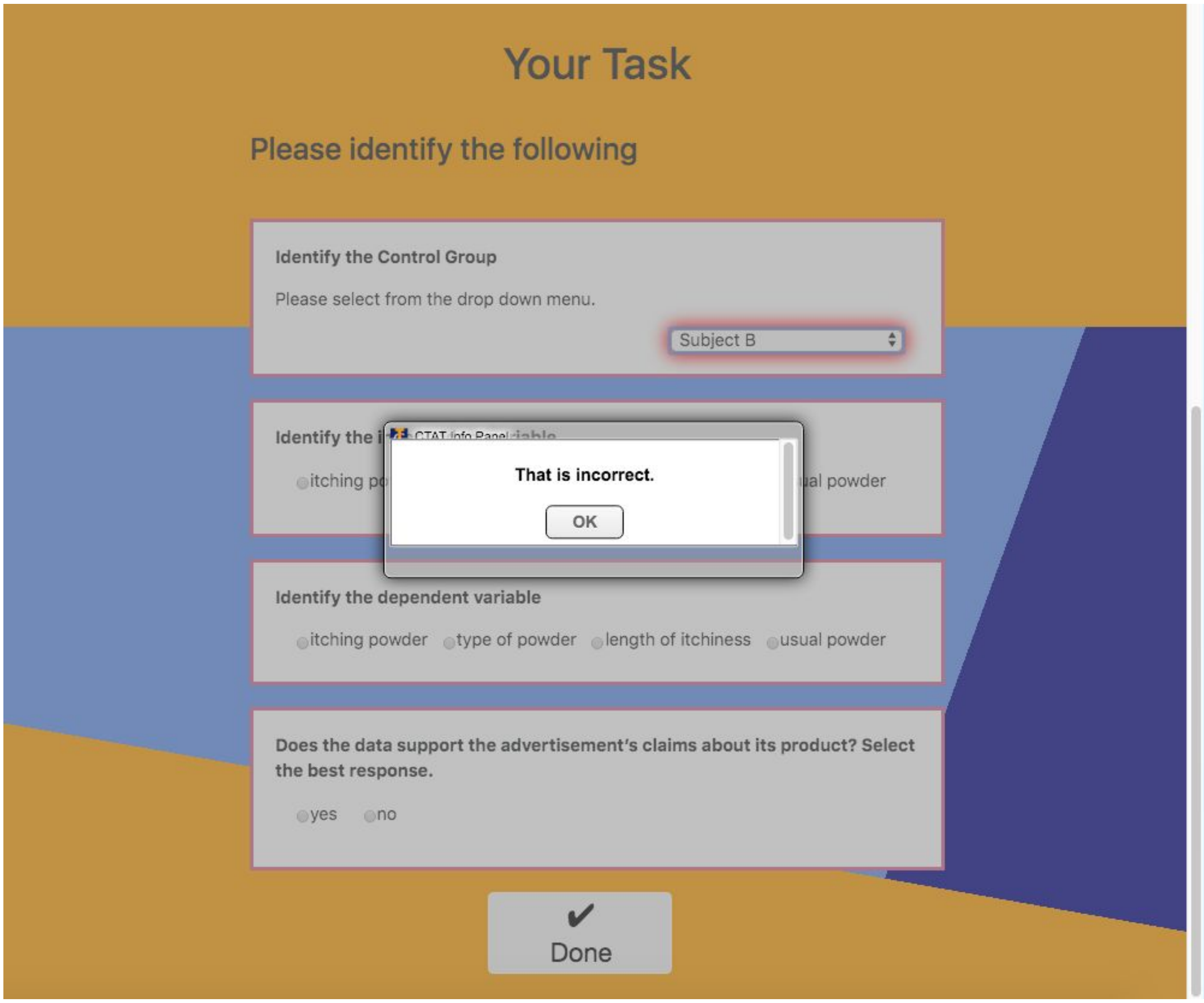


Figure 3. Feedback Condition 2

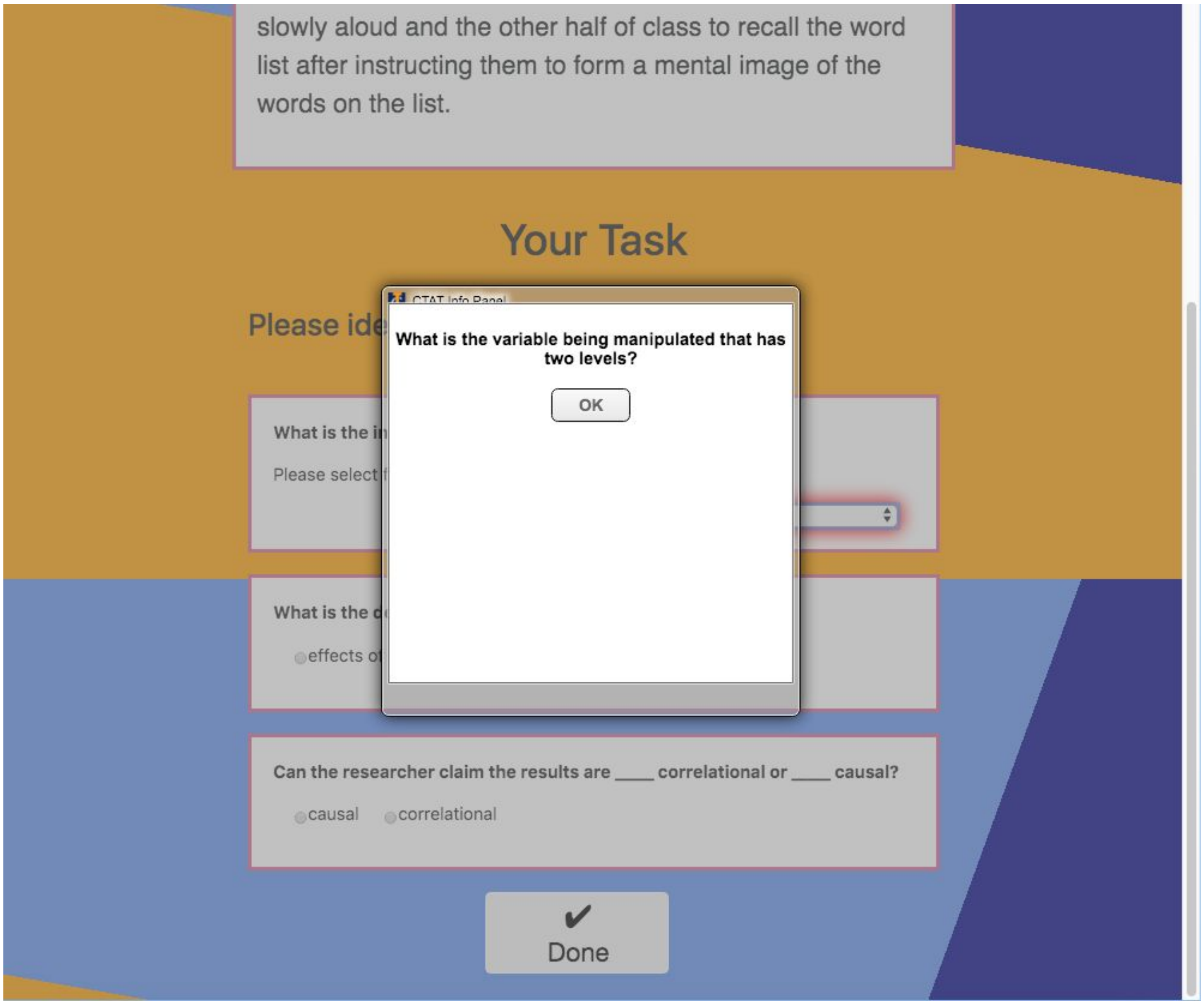


Figure 4. Feedback Condition 3

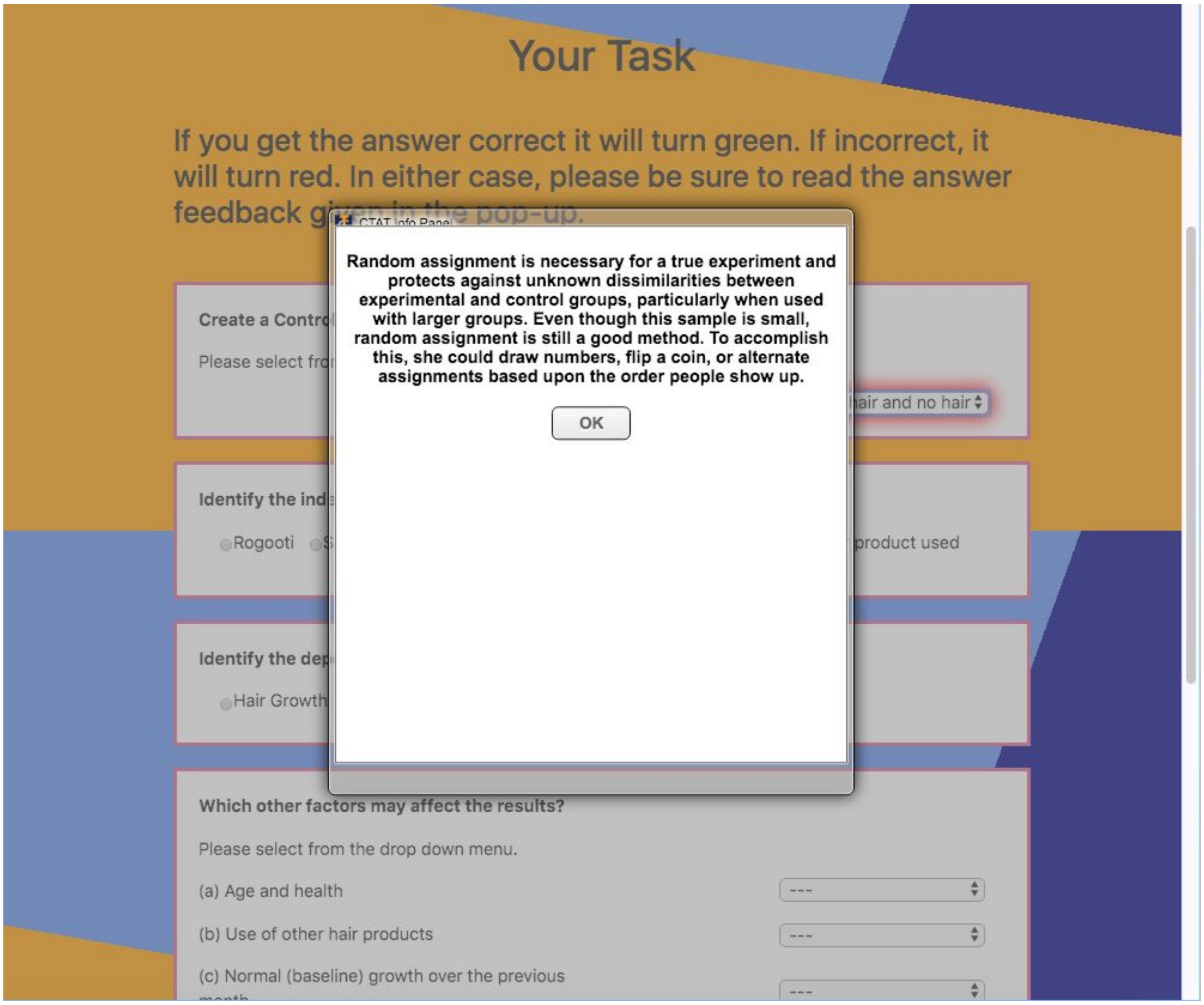


Figure 5.

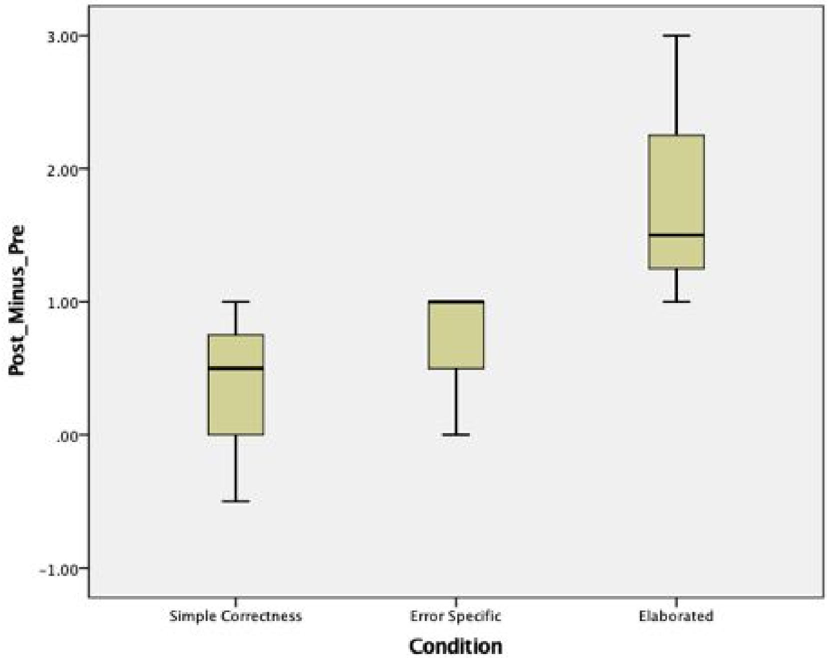


Figure 6.

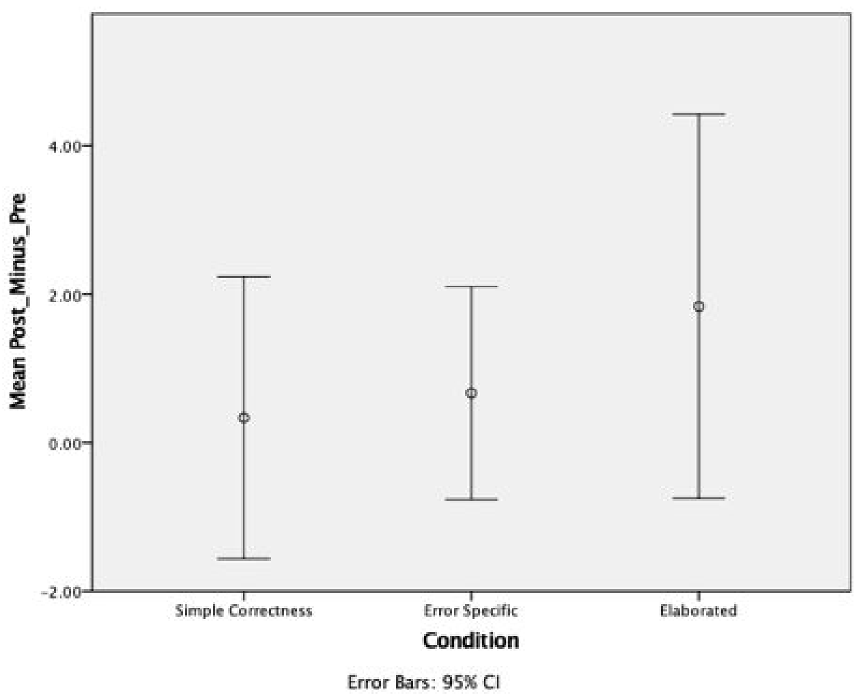


Figure 7.

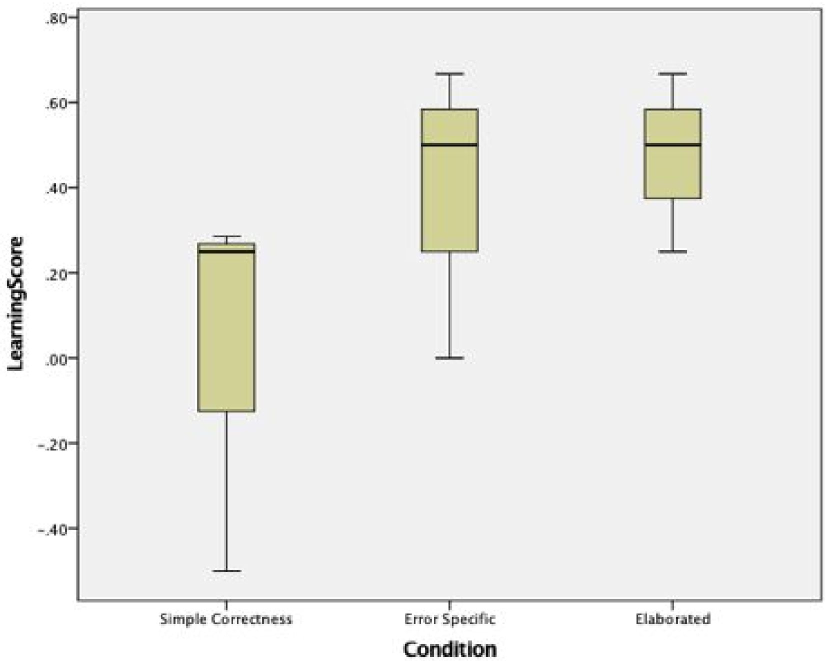
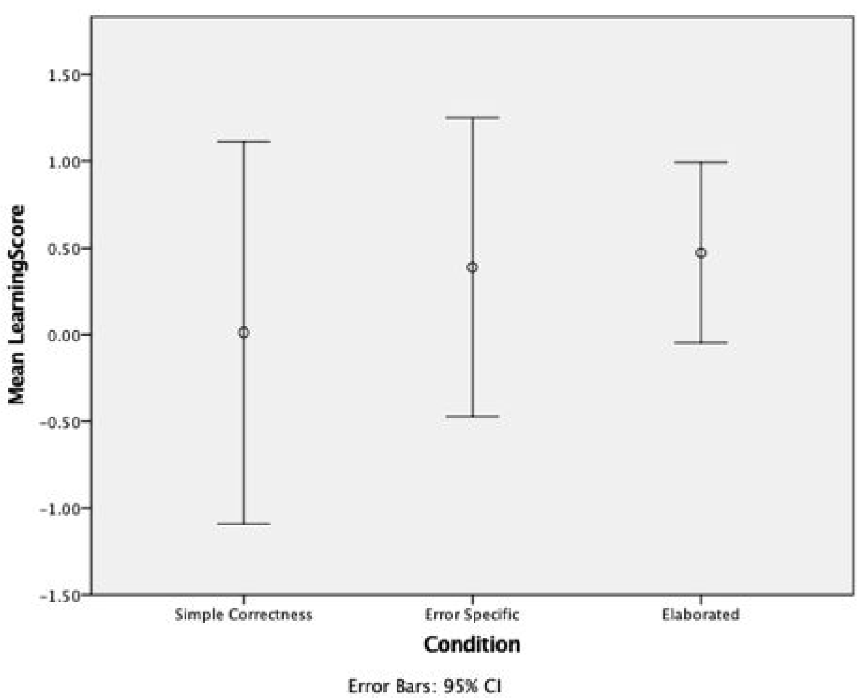


Figure 8.



Instructions on How To Access TutorShop Tutors

1. Link to TutorShop: https://school.tutorshop.web.cmu.edu/
    a. Username: pcasula
    b. Password: pcasula
2. Packages → Pol2 Final Tutor → Problem Sets
    a. Tutor 1 : Simple Correctness Feedback condition
    b. Tutor 2: Error-Specific Feedback condition
    c. Tutor 3: Elaborated Feedback condition

References

1. Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4*(2), 167-207.

2. Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112

3. McKendree, J. (1990). Effective feedback content for tutoring complex skills. *Human-Computer Interaction, 5*(4), 381-413. doi:10.1207/s15327051hci0504_2

4. Sleeman, D., Kelly, E., Martinak, R., Ward, D., & Moore, L. (1989). Studies of diagnosis and remediation with high school algebra students. *Cognitive Science*, *13*(4), 551-568. doi:10.1207/s15516709cog1304_3

5. Van der Kleij, F. M., Feskens, R. C., & Eggen, T. J. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research, 85*(4), 475-511.