Response Justifications as Feedback in Clicker Activities: A Case Study on Student Performance and Calibration

Pantelis M. Papadopoulos[1], Antonis Natsis[1], Nikolaus Obwegeser[2]
[1]Centre for Teaching Development and Digital Media, Aarhus University, Aarhus, Denmark
[2]Department of Management, Aarhus University, Aarhus, Denmark
pmpapad@tdm.au.dk, anatsis@tdm.au.dk, nikolaus@mgmt.au.dk

Abstract: The study examines the potential of short justifications in clicker activities. A total of 138 students answered individually eight multiple-choice questions in a clicker tool and provided short justifications for their responses, denoting also their confidence that their responses were correct. Next, students received classroom feedback and revised their initial answers and their level of confidence. Results showed that all students increased their performance during the revision phase. However, the group ($n = 70$) that received as feedback the percentage of students under each question choice along with the respective justifications increased its confidence significantly. Moreover, in this group, students' final confidence levels and their actual performance were significantly positively correlated, suggesting accuracy between their perceived and actual performance (i.e., better calibration). On the contrary, the same was not observed for the group ($n = 68$) that received as feedback only the percentage information. This suggests that having access to justifications helps students in self-assessing their level of knowledge more accurately.

Introduction

Higher learning outcomes seem to be associated with metacognitively aware students (Mayer, 2008). Therefore, the development of learning environments that support students in promoting their metacognitive knowledge and skills is of high significance. These learning environments set the ground for students to control and reflect upon their learning. In this context, students are able to self-monitor their progress, adjust their efforts as well as, select and implement suitable learning strategies to reach the desired learning goals (Efklides, 2006). The selection of proper instructional methods by the teachers as well as the effective use of technology can further facilitate the promotion of students' metacognition.

In this view, clickers (a.k.a. student response systems, audience response systems, etc.) is one specific technology that has gained increasing attention over the recent years since their use can contribute to cognitive and affective learning gains for the students (Hunsu, Adesope, & Bayly, 2016). Moreover, there are indications that their use can positively influence students' metacognition (Brady, Seli, & Rosenthal, 2013) due to the high interaction among students, peers and teachers and also, due to the provided feedback. Clickers allow teachers to ask students questions, which often take the form of a multiple-choice quiz, at various times within a lecture. Students respond, often anonymously, and they can get directly customized feedback on their responses and the aggregated responses of their fellow students. Therefore, students are given the opportunity to self-evaluate the degree of their own comprehension (Mayer et al., 2009) and compare it with their peers (Ioannou & Artino, 2010) resulting in the enhancement of their self-monitoring of learning. The tallied answers can also be presented to the teachers allowing them to identify students' misconceptions or identify concepts that need clarifications or organize discussions among the students based on the received feedback.

The most common implementation of clickers is based on the peer instruction method (Mazur, 1997, 2009). Firstly, students provide their individual answers (vote) to a multiple-choice question and then get feedback regarding student population for each selected option. Following a short discussion session with their peers, they answer again (revote) and view updated response results. Lastly, they get informed on the correct answers followed by teacher's explanations or class wide discussions. As various studies have demonstrated (e.g. Crouch, Watkins, Fagen, & Mazur, 2007; Zingaro, & Porter, 2014), the successful implementation of the peer instruction method can improve various aspects of the learning process and lead to greater learning benefits for students compared to a traditional lecture.

Yet these benefits are fewer for students who are not keen on social interaction or hesitate to exchange their ideas when discussing with their fellow students (Michinov, Morice, & Ferrières, 2015). Furthermore, the use of clickers in the classroom is considered by some educators as a time-consuming activity (Lanz & Stawiski, 2014) that can interrupt the smooth flow of the course (Katz, Hallam, Duvall, & Polsky, 2017), since it is difficult to handle variations of lecturer input and discussion, to provide appropriate feedback and guarantee students' active involvement (Nicol & Boyle, 2003). Moreover, the feedback on clickers, that is used to trigger peer discussions, is based primarily on the population (i.e., absolute number or percentage) of students that selected each question choice. This type of feedback is usually presented to students through figures or graphs. We argue that this kind of information, helpful as it might be to students, cannot provide sufficient insight on their peers. As such, students would be benefited, particularly in self-assessing and comparing themselves with their fellow students, if feedback included information that could better describe the clicker audience.

In the current study, a modified version of the use of clickers with the peer instruction method is presented, after taking into consideration the aforementioned issues. In the first phase, students answer multiple-choice questions individually, providing also a short justification on their selection and their confidence on whether their answers were correct. Asking students to provide written answers in order to make their thinking explicit may result in enhanced outcomes in technology-enhanced learning environments (Papadopoulos, Dimitriadis, Stamelos, Tsoukalas, 2011). Through self-explanation, students are prone to identify possible knowledge gaps and to enhance their metacognitive processes of detecting and correcting errors and so, reach a deeper understanding of learning materials (Chi, & VanLehn, 2010). When these explanations are presented to their peers, they may lead to social interactions. In this context, students may consider their peers' explanations as external input and use them to control and coordinate their learning with reference of the group, which they are involved in (Chien, Chang, & Chang, 2016). Therefore, in the present study, students did not engage in peer discussions before answering for the second time the teacher's questions. Instead, they employed different feedback information including the class responses and their peers' justifications.

Additionally, in the first phase, students were asked to judge how confident they felt about the correctness of their answer. Confidence judgements illustrate students' awareness of their own learning and as such, depict their metacognitive monitoring skills (Hadwin, & Webster, 2013). Through the difference in students' self-reported level of confidence and their actual performance on the questions, calibration accuracy can be measured (Gutierrez & Schraw, 2015). Calibration is the level of agreement between students' perceptions of their performance and their actual performance in an assessment task (Nietfeld, Cao, & Osborne, 2006). That, in turn, is important, from a metacognitive perspective, because accurate calibration enables students to better monitor their progress, adapt their learning efforts and behavior and consequently, self-regulate their own learning (Alexander, 2013).

Based on the above, the current study aims at presenting preliminary analysis on the impact of integrating short justifications as feedback on clicker activities. More specifically, the research question that the present study aims to examine is the following: What is the impact of integrating peer justifications as feedback on students' performance, confidence and calibration?

Method

Participants and Domain

The course "Business Development with Information Systems" is a 5-ECTS Bachelor course, offered typically in the third semester in the Department of Management. The lecture material (i.e., slides, literature, etc.) is available online a week in advance. As a common practice, students are urged and expected to be knowledgeable on the material before coming into the class. However, this task is not mandatory. Lectures are given weekly in a large auditorium and last 2 hours.

A total of 138 students volunteered to participate in the study that was not part of the formal course assessment. Students were randomly distributed by the system into two treatment conditions, according to the type of feedback information they received during the revision phase in the clicker activity:

- Percentage (PERC): Students that received only percentage information as feedback (n = 68).
- Percentage and Justifications (PERC_JUST): Students that received both percentage information and peer justifications as feedback (n = 70).

The SAGA Tool

The "Self-Assessment/Group Awareness – SAGA" tool is a web-based, cross-platform clicker tool that was designed and developed by our university and has been used in previous studies on clickers (Papadopoulos, Natsis,

& Obwegeser, 2017). The tool takes the students through two answering rounds, similar to the vote-revote phases in the peer instruction paradigm (Mazur, 1997, 2009). However, a characteristic that differentiates SAGA from most of the other clickers is the inclusion of additional feedback information on peer activity during the second round (which, in turn, includes a revision, instead of a re-voting task).

During the first phase, the students answer individually a set of eight teacher-generated multiple-choice questions with four choices each. For each question, students also have to submit their level of confidence that their answer is correct (using a 5-step Likert scale) and add short justifications (i.e., 140 characters) on why they think their choice is the correct one. The first phase is common for all students in the tool. In the second phase, the students see their previous answers, their level of confidence, the percentage of their peers in the classroom that selected each question choice, and, depending on their feedback condition, the justifications their peers wrote under each choice. Then, the students are able to revise their initial answers and the respective levels of confidence for each question (Figure 1).

Figure 1. Screenshot of the SAGA tool during the revision phase (PERC_JUST).

After the completion of the revision phase, the students are able to see their scores and the correct answers and discuss them with the teacher. All students are in the same phase simultaneously during the activity and SAGA provides monitoring functionalities to the teacher, who is responsible for activating the next phase in the process.

Procedure and Study Conditions

The students were informed about the research nature of the activity and the fact that their fellow students may receive different information from the system. The activity was conducted in the first 20 minutes of a lecture. The students were allowed 10 minutes to provide their initial answers. Immediately after that, students had 5 minutes to see the feedback (which was different in the PERC and PERC_JUST groups), and revise their initial answers. During the last five minutes, the students saw their scores and the correct answers, while the teacher provided additional explanations. After that, the planned lecture started. The whole activity was individual and anonymous. No personal information about the students was recorded by SAGA or the teacher.

Research Design and Data Analysis

The study employed a between-subjects research design with the type of feedback each group received being the independent variable and students' performance, confidence levels, and justifications being the dependent ones. For all statistical analyses, a level of significance at .05 was chosen. Parametric tests were used in data analysis, since no test assumption was violated.

Results

Table 1 presents student activity during the initial and the revision phase of the activity, in terms of performance, confidence, and justifications. T-test results showed that the two groups were comparable during the first phase, having similar initial performance, confidence, and justification length (in characters) ($p > 0.05$). In addition, we compared the mean length of justification between cases in which students had selected the correct and an incorrect question choice. T-test analysis revealed no significant difference, suggesting that students had written justification of comparable length regardless of the correctness of their choice ($p > 0.05$).

Regarding the revision phase, one-way analysis of covariance (one-way ANCOVA) results revealed no significant effect of the treatment condition on the revised performance, after controlling for initial performance ($p > 0.05$). However, one-way ANCOVA results showed that there was a significant difference in the revised confidence score for the two groups, after controlling for initial confidence ($F(1,135) = 4.305$, $p = 0.040$, $\eta^2 = 0.031$), with the PERC_JUST group feeling more confident.

Table 1: Student activity

	Percentage			**Percentage & Justifications**		
PERFORMANCE (0-8)	**M**	***SD***	**n**	**M**	***SD***	**n**
Initial performance	3.92	*(1.26)*	68	5.20	*(1.67)*	70
Revised performance	5.36	*(1.73)*	68	6.66	*(1.93)*	70
CONFIDENCE (1-5)						
Initial confidence	2.75	*(1.00)*	68	2.76	*(0.86)*	70
Revised confidence	2.94	*(1.01)*	68	3.12	*(0.86)*	70
JUSTIFICATIONS (1-140)						
Justification length (char.)	24.41	*(20.71)*	68	28.84	*(21.67)*	70

Pearson's bivariate correlation coefficient results revealed a significant positive correlation between the length of the justifications the students provided and their initial confidence ($r = 0.264$, $n = 138$, $p = 0.002$), suggesting that students that felt more confident, also wrote longer justifications for their answers. Similarly, initial and revised performance scores were also positively correlated ($r = 0.838$, $n = 138$, $p < 0.001$).

Paired-samples t-test results showed that students in both treatment groups significantly increased their performance and their confidence during the revision phase. However, bivariate correlation analysis in the two treatment groups showed that students' confidence and performance scores were positively correlated in the revision phase only in the PERC_JUST group ($r = 0.227$, $n = 70$, $p = 0.048$), while no such correlation occurred in the PERC group. This suggests that the confidence level of students that received both percentage and peer justifications as feedback was aligned to their actual performance.

Discussion and Conclusions

Result analysis showed that all students were able to improve their initial performance during the revision phase of the activity. However, analysis also showed that the addition of the justification information was not enough for the students in the PERC_JUST condition to outperform students that only received percentage feedback. Nevertheless, this enriched feedback was enough to produce a significant difference regarding the confidence levels between the two treatment groups in the revision phase. It seems that providing students with their peer explanations as feedback allowed them to effectively use them to monitor and regulate their learning progress. Furthermore, our results are aligned with current literature regarding the significance of asking students about their confidence when answering questions (e.g., Kleitman & Costa, 2014; Schnaubert & Bodemer, 2015). One explanation for the increased confidence of the PERC_JUST group is that peer justifications added another layer of reassurance for students. While the percentage information presented the size of student populations under each question choice, the justifications presented the voices and perspectives of these populations. In other words, students were able to get a better view on why their peers thought a choice was correct and this information either re-enforced their initial views or persuaded them to revise their answers. In both cases, the students emerged from the revision phase feeling significantly more confident than their counterparts that received only feedback based on percentage. We argue that for the latter, engaging in revisions knowing only which choice the others selected may also create frustration, especially in cases in which a student's answer goes against the classroom majority.

In addition, correlation analysis showed that the increased confidence in the PERC_JUST group was in line with the increased performance, while no similar finding was observed in the PERC group. This suggests a better calibration between the perceived (i.e., level of confidence) and the actual performance for students that had access to peer justification in the revision phase. In other words, the lack of justifications in the PERC group caused either false confidence or unjustified uncertainty. Thus, we argue that although the examined instructional intervention did not result in higher student performance, it still offered learning gains, since it improved their calibration. Calibration, as a facet of metacognition, is crucial for students, since it can be linked to increased academic performance, while enhancing cognitive and metacognitive skills (Alexander, 2013). We maintain that in a longer treatment period, this improved calibration could enhance self-efficacy and, ultimately, higher acquisition of domain knowledge.

Finally, without suggesting that the length of a justification is necessarily linked to the quality of the argument presented, results showed that students that were certain about their answers also provided longer justifications. However, as correlation analysis also suggested, this high confidence was not always based on correct responses. Analysis on the actual quality of students' justification could have provided a more accurate picture on student activity. However, it was not feasible at this point to review the over 1,000 justifications submitted and this activity is planned for a future analysis. Moreover, this future analysis will focus on the most ambiguous questions, i.e., questions in which students seemed divided between at least two choices during the initial phase).

In conclusion, this study provided empirical evidence on the potential of student-generated short justifications as addition feedback information, alongside the commonly used percentage metric, in clicker tools. The results showed that such information could increase student confidence and improve calibration, while a longer study design could also identify a main effect on students' performance.

Acknowledgement

This work has been partially funded by a Starting Grant from AUFF (Aarhus Universitets Forskningsfond), titled "Innovative and Emerging Technologies in Education".

References

Alexander, P. A. (2013). Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning and Instruction, 24,* 1-3.

Brady, M., Seli, H., & Rosenthal, J. (2013). "Clickers" and metacognition: A quasi-experimental comparative study about metacognitive self-regulation and use of electronic feedback devices. *Computers & Education, 65*, 56-63.

Chi, M., & VanLehn, K. (2010). Meta-Cognitive Strategy Instruction in Intelligent Tutoring Systems: How, When, and Why. *Journal of Educational Technology & Society, 13*(1), 25-39.

Chien, Y.-T., Chang, Y.-H., & Chang, C.-Y. (2016). Do we click in the right way? A meta-analytic review of clicker-integrated instruction. *Educational Research Review, 17*, 1-18.

Crouch, C. H., Watkins, J., Fagen, A. P., & Mazur, E. (2007). Peer instruction: Engaging students one-on-one, all at once. *Research-Based Reform of University Physics, 1*(1), 40-95.

Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review, 1*(1), 3-14.

Gutierrez, A. P., & Schraw, G. (2015). Effects of Strategy Training and Incentives on Students' Performance, Confidence, and Calibration. *The Journal of Experimental Education, 83*(3), 386-404.

Hadwin, A. F., & Webster, E. A. (2013). Calibration in goal setting: Examining the nature of judgments of confidence. *Learning and Instruction, 24*, 37-47.

Hunsu, N. J., Adesope, O., & Bayly, D. J. (2016). A meta-analysis of the effects of audience response systems (clicker-based technologies) on cognition and affect. *Computers & Education*, *94*, 102-119.

Ioannou, A., & Artino, A. R. (2010). Using a classroom response system to support active learning in an educational psychology course: A case study. *International Journal of Instructional Media, 37*(3), 315-325.

Katz, L., Hallam, M. C., Duvall, M. M., & Polsky, Z. (2017). Considerations for using personal Wi-Fi enabled devices as “clickers” in a large university class. *Active Learning in Higher Education, 18*(1), 25-35.

Kleitman, S., & Costa, D. S. J. (2014). The role of a novel formative assessment tool (Stats-mIQ) and individual differences in real-life academic performance. *Learning and Individual Differences, 29*, 150-161.

Lantz, M. E., & Stawiski, A. (2014). Effectiveness of clickers: Effect of feedback and the timing of questions on learning. *Computers in Human Behavior, 31*, 280-286.

Mayer, R. E. (2008). *Learning and Instruction*: Pearson Merrill Prentice Hall.

Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., et al. (2009). Clickers in college classrooms: fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology*, *34*(1), 51-57.

Mazur, E. (1997). *Peer instruction: A user's manual series in educational innovation*. Upper Saddle River, NJ: Prentice Hall.

Mazur, E. (2009). Farewell, lecture? *Science, 323*, 50-51.

Michinov, N., Morice, J., & Ferrières, V. (2015). A step further in Peer Instruction: Using the Stepladder technique to improve learning. *Computers & Education, 91*, 1-13.

Nicol, D. J., & Boyle, J. T. (2003). Peer Instruction versus Class-wide Discussion in Large Classes: A comparison of two interaction methods in the wired classroom. *Studies in Higher Education*, *28*(4), 457-473.

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition and Learning, 1*(2), 159-179.

Papadopoulos, P. M., Demetriadis, S. N., Stamelos, I. G., & Tsoukalas, I. A. (2011). The value of writing-to-learn when using question prompts to support web-based learning in ill-structured domains. *Educational Technology Research and Development, 59*(1), 71-90.

Papadopoulos, P.M., Natsis, A., & Obwegeser, N. (2017). Improving the Quiz - Student Preparation and Confidence as Feedback Metrics.In P. Escudeiro, G. Costagliola, S. Zvacek, J. Uhomoibhi, & B.M. McLaren (Eds.), *Proceedings of the 9th International Conference on Computer Supported (CSEDU)*, (pp. 59-69).

Schnaubert, L., & Bodemer, D. (2015). Subjective Validity Ratings to Support Shared Knowledge Construction in CSCL. In O. Lindwall, P. Häkkinen, T. Koschmann, P. Tchounikine, & S. Ludvigsen (Eds.), *Exploring the Material Conditions of Learning: The Computer Supported Collaborative Learning (CSCL) Conference 2015 (Vol. 2)* (pp. 933-934). Gothenburg: International Society of the Learning Sciences.

Zingaro, D., & Porter, L. (2014). Peer Instruction in computing: The value of instructor intervention. *Computers & Education, 71*, 87-96.