Journal of Computer Assisted Learning

Enriching Feedback in Audience Response Systems: Analysis and Implications of Objective and Subjective Metrics on Students' Performance and Attitudes

Journal:	*Journal of Computer Assisted Learning*
Manuscript ID	JCAL-18-214.R1
Manuscript Type:	Article
Technology and Tools:	Groupware, Hypertext/Hypermedia, Information systems
Learning process / Pedagogy:	Collaborative Learning, Individual learning, Peer Learning
Paradigm:	Constructivism
Level of education :	Undergraduate
Place of learning:	Formal Learning
Type of research:	Quantitative, Mixed methods
Research technique:	Experimental, Case Study
Analysis/evaluation paradigm:	Structural Equation Modelling
Issues:	Knowledge, Attitude, Satisfaction

What is currently known about the subject matter

- Audience response systems (ARS) typically provide anonymous feedback (i.e., percentage)
- Information on how much a partner knows is important for peer answer evaluation
- Systematic research of such metacognitive information on peers in ARS is missing

What their paper adds to this

- This study explores the use of objective and subjective feedback metrics in ARS
- These metrics are: percentage, preparation, confidence, and past performance
- It also analyzes how such metrics may impact student attitudes and performance

Implications of study findings for practitioners

- Main effect for preparation and confidence metrics in "challenging" questions
- Positive student evaluation of ARS for deeper understanding and clarifying misconceptions

Enriching Feedback in Audience Response Systems: Analysis and Implications of Objective and Subjective Metrics on Students' Performance and Attitudes

Abstract: The aim of the present study (n = 113) was to examine how (objective and subjective) information on peers' preparation, confidence, and past performance can support students in answering correctly in audience response systems (ARS; aka clickers). The result analysis shows that in the "challenging" questions, in which answers diverged, students who received additional information about peers' self-reported preparation and/or confidence outperformed students who were only given the objective percentage with or without past performance feedback. In addition, students expressed a positive attitude towards the activity, commenting its usefulness in better understanding course material and identifying misconceptions.

Keywords: Audience Response System, Clickers, Feedback, Confidence, Preparation, Formative Assessment.

1. The Role of Peers in Audience Response Systems

Asking challenging questions in the classroom could help students identify misconceptions and better understand the course material. One downside, or so it may seem, could lie in the fact that teachers can only call upon one student at a time to answer the question, while any other student may digress off-topic. Audience response systems (ARS; aka clickers) allow the teacher to pose, usually closed-type, questions to each student of a class at once, aggregate received answers, and provide immediate, personalized feedback to the students (Chien, Chang, & Chang, 2016). The feedback that such systems typically provide is anonymous, which can enhance

psychological safety and acceptance of quizzes within the classroom (Barr, 2017; Bojinova & Oigara, 2013; Stowell et al., 2010), but may also lack important information about the peers. While ARS addressing all students at once may reduce the problem of student disengagement in the classroom (Gehlen-Baum, Weinberger, Pohl, & Bry, 2012), interacting via ARS may disable observing and building on the reasoning of peers. Orientation towards known well-performing and confident students serving as models seems to be important for feedback to take effect, not only on single students, but the whole classroom (Drabman & Lahey, 1974, Smith et al., 2009). While students receive this information, to a certain degree, when they discuss their answers in small groups, ARS ability to present information on the whole class may be an advantage. Consequently, studies on online assessment and group awareness have argued in favor of students' level of confidence as a useful feedback metric to be included in ARS (e.g., Kleitman & Costa, 2014; Schnaubert & Bodemer, 2015). Focusing on cognitive group awareness tools, Engelmann et al. (2009) proposed a dichotomy between tools that provide information on peers' knowledge (Dehler, Bodemer, Buder, & Hesse, 2011) and tools that provide metacognitive information on how much a partner knows (e.g., Erkens & Bodemer, 2015; Sangin, Molinari, Nüssli, & Dillenbourg, 2011). The latter require the learners to explicitly evaluate their knowledge, thus producing subjective metrics (Schnaubert & Bodemer, 2015) whose usefulness (i.e., how indicative they are to be able to support students' decision-making) and validity (i.e., accurate depiction of information) depend, in turn, on learners' self-assessing skills.

So here, we aim at investigating the combination of question asking and different metacognitive information on the characteristics of students selecting each answer choice to leverage the potential of ARS to address all students at once and

simultaneously foster using peers as additional resource. Vickrey et al. (2015) provided an overview of factors that could affect question asking under the Peer Instruction paradigm (Mazur, 1997) mentioning also a possible relationship between learning gains observed and individual student characteristics. Nevertheless, there is yet little research on using peer metacognitive information as feedback in ARS. Feedback in ARS typically builds on the number or percentage of students that selected each question choice. This percentage information of answer choice, however, may encourage students to focus more on probabilistic reasoning shifting their initial answers to the most popular one (Nielsen et al., 2012; Perez et al., 2010). Therefore, we expect that enriching feedback with metrics providing a better picture of the peers would be beneficial for students, especially in self-assessment and in comparing themselves with their peers.

This study (a) explores the use of complementary feedback metrics that additionally provides information about the peers in an ARS activity alongside the percentage information of answer choices made, and (b) analyzes how such metrics may impact student attitudes and performance. The three additional metrics examined presented information on:

- how prepared the students were feeling before starting the activity (preparation metric),
- how confident they were feeling after answering a question (confidence metric), and
- how they had performed in previous ARS activities (past performance metric).

The selection of these two subjective (i.e., preparation and confidence) and one objective metrics (past performance) was based on the assumption that the additional feedback presented may be useful for students in questions where students' answers diverge and the percentage information is ambiguous.

2. What is Known about ARS?

2.1 Learning Benefits of ARS

The use of ARS has been repeatedly linked to increased engagement and motivation (Caldwell, 2007; Crouch & Mazur, 2001; Siau, Sheng, & Nah, 2006). Specifically, studies have reported that students that use such systems pay more attention during the class (Gibbs & Simpson, 2004), have higher attendance rates (Poirier & Feldman, 2007), are more likely to ask or answer questions during a lecture (Caldwell, 2007; Wit, 2003), and they find their classes more enjoyable (Ioannou & Artino, 2010) and satisfying (Marshall, Valdosta, & Varnon, 2012).

Students engaged in ARS activities are invested in their answers and they are more prone to participate in group and/or classroom discussions (Fies & Marshall, 2006; Nicol & Boyle, 2003). Furthermore, clickers are routinely credited for developing critical thinking (Mollborn & Hoekstra, 2010) uncovering preconceptions and assumptions (Hoekstra & Mollborn, 2012), and monitoring students' progress (Caldwell, 2007). In addition, being unable to cross barriers often created by teachers' attitudes towards technology (Blackwell, Lauricella, Wartella, Robb, & Schomburg, 2013; Ertmer, Ottenbreit-Leftwich, Sadik, Sendurur, & Sendurur, 2012), studies have repeatedly noted teachers' appreciation towards audience response systems in

clarifying concepts, identifying misconceptions, and organizing lectures (Caldwell, 2007; Chen, Whittinghill, & Kadlowec, 2010; Fies & Marshall, 2006).

Despite their recognized educational value, reviews on ARS present an inconclusive view on their potential in enhancing student learning and academic performance. On one side, numerous studies have provided evidence on their beneficial impact on retention (Prince, 2004), factual knowledge acquisition (Mayer et al., 2009; Shapiro & Gordon, 2012, 2013; Shapiro et al., 2017), and course grades (Brady, Seli, & Rosenthal, 2013; Mayer et al., 2009; Poirier & Feldman, 2007). For instance, Mayer et al. (2009) compared the midterm and final exam scores of students that answered clicker questions, students that answered the same questions on paper, and students that did not receive questions. Students in the ARS condition outperformed the other two groups, suggesting that it was the tool and not the questions alone that resulted in higher scores.

On the other side, however, there are also studies showing that non-clicker classes were comparable (Caldwell 2007; Elicker & McConnell, 2011) or even better than the clicker ones (Fortner-Wood et al., 2013). Although Shapiro (2009) cites methodological issues and technology-related factors as the reason for little or negative effects in some studies, the fact remains that the discussion on the impact of ARS on actual student performance is still ongoing. Thus, many researchers have called for further studies that would integrate both qualitative and quantitative data in rigorous research designs (Chien et al., 2016; Fies & Marshall, 2006; Han, 2014; Kay & LeSage, 2009), while others underlined the importance of the instructional approach (Kennedy & Cutts, 2005; MacGeorge et al., 2008; Mun, Hew, & Cheung, 2009; Shapiro et al., 2017).

2.2 Peer Interaction and Reflection with ARS

There is a rich literature on instructional approaches around ARS tools (for an overview, see Caldwell, 2007; Chien et al., 2016; Fies & Marshall, 2006; Han, 2014; Hunsu, Adesope, & Bayly, 2016; Kay & LeSage, 2009). Clicker-based activities can be designed as individual or collaborative activities (McDonough & Foote, 2015), utilize different forms of closed-type questions, such as multiple-choice, yes/no, or fill-in-the-blank items (Desrochers & Shelnutt, 2012; Sutherlin, Sutherlin, & Akpanudo, 2013), and be used at different points during a lecture for addressing different sets of learning goals. For example, using clickers in the beginning or during the lecture could provide valuable information on students' prior understandings and misconceptions (Anderson, Healy, Kole, & Bourne, 2013; Caldwell, 2007), thus helping the teacher in better adapting the lecture to students' needs (Kay & LeSage, 2009; Wit, 2003). Then again, using clickers at the end of the lecture could provide opportunities for reflection and enhance retention (Dufresne, Gerace, Leonard, Mestre, & Wenk, 1996).

Arguably, most of the pedagogical approaches around ARS found in the literature are based, at least partially, on the Peer Instruction method suggested by Mazur and his colleagues (e.g., Crouch & Mazur, 2001; Mazur, 1997, 2009). The scope of Peer Instruction is to increase engagement in the classroom and support student discussion and reflection on multiple perspectives and ideas. The method adopts elements of the think-pair-share collaboration script, engaging students in individual reflection, before sharing their understandings with a few peers and, later on, the class (Watkins & Mazur, 2010). Peer Instruction employs a structured questioning process that is usually organized as follows:

1. Students answer (vote) individually a multiple-choice question, using an audience response system.
2. Feedback in the form of tallied voting results is presented to the students by the ARS.
3. If there is a low percentage of correct answers, the teacher asks the students to discuss briefly their answers with their neighbors.
4. The students answer the same question for a second time (revote).
5. Finally, the students receive corrective feedback and engage in a class discussion, during which the teacher offers additional explanations.

The re-voting phase is important, because it invites the students to reflect and act upon the feedback they received from the tool (i.e., distribution of student population under each question choice) and their peers (i.e., sharing ideas during the brief discussion), thus closing the "feedback loop" (Boud & Molloy, 2013). In addition, past studies have examined the impact of presenting the tally of initial answers before or after peer discussion (e.g., Brooks & Koretsky, 2011; Nielsen et al., 2012; Perez et al., 2010), with Vickrey et al. (2015) suggesting that presenting the tally of initial answers after peer discussion may limit bias towards the consensus, while also mentioning that more research is needed.

The method has been repeatedly associated with better student performance (e.g., Crouch, Watkins, Fagen, & Mazur, 2007; Crouch & Mazur, 2001; Mayer et al., 2009) and increased interaction between students and teachers and between peers (Blasco-Arcas, Buil, Hernández-Ortega, & Sese, 2013). At the same time, a plethora or variants and extensions can be found in the literature (for an overview, Vickey et al., 2015). Michinov, Morice, and Ferrières (2015), for example, addressed the often observed unwillingness of students' to engage in fruitful peer discussions during

clicker activities, by proposing the integration of the stepladder technique in the Peer Instruction approach. According to this technique, new members enter a student group sequentially and are forced to participate to the discussion. Mazur (2009) pointed out the importance of evaluating ARS tools in conjunction with the educational context and the underlying pedagogy, suggesting that over the years the "clicker method" term has been used to refer to educational settings distant to the Peer Instruction approach he introduced, stating that "it is not the technology, but the pedagogy that matters" (2009, p.51).

2.3 Study Motivation and Research Questions

Feedback has been found to be among the most effective components within a learning environment as it provides additional, personalized learning opportunities for exactly those contents learners struggle with (e.g., Hattie, 2009; Hattie & Timperley, 2007). Whereas confirmative feedback has been considered from a reinforcement perspective, corrective feedback is supposed to lead to cognitive elaboration and correction of mental schemes. Here, informative feedback outperforms simple grading or feedback on correctness, providing a basis for modification of existing knowledge structures and schemes. Interestingly, certitude of learners' responses can interact with corrective feedback in that learners confidently giving wrong answers and being corrected invest more mental effort in studying the feedback and subsequently attaining higher learning gains (Griffiths & Highman, 2018; Hancock, Stock, & Kulhavy, 1992; Kulhavy & Stock, 1989).

Yet, little is known in classroom research as to how "confidence by proxy" (i.e., relying on someone else's confidence) influences feedback effects on the classroom

level. Learners do not only judge their own certitude, but also the certitude of their peers that they witness in class. For example, a student that is known by peers to be above average knowledgeable may serve as a model, the student's responses will be considered as probably correct, and respectively peers may be more likely to adjust their understanding and imagined responses to the ones given by the respective "good" student. While students typically have a good understanding about the knowledge levels of their peers, an anonymous ARS response matrix fails to provide information on certitude and prior performance of the responders. Instead, a different sort of confidence on the majority vote can be deduced in what has been considered wisdom-of-the-crowd information. Adding data on certitude and past performance to the response matrix can serve to investigate how this additional information influences learners' consecutive responses, which would not be possible without technology support providing anonymous information of what responses peers chose.

We followed Mazur's Peer Instruction paradigm (e.g., Crouch & Mazur, 2001; Mazur, 1997, 2009), substituting the brief peer discussion session with enriched feedback that included, in addition to percentage, the preparation, confidence, and past performance mentioned earlier. The reason for this departure from Mazur's paradigm was twofold: adhering to setting constraints and focusing on the impact of feedback on students' answering strategies. Regarding constraints, the activity had to be short, not to disrupt the lecture flow (a common issue during clickers implementation – e.g., Koenig, 2010; Strasser, 2010), while the large audience (>100) and the auditorium in which the activity took place were also factors against peer discussion. Moreover, asking students to revisit the questions based only on the provided feedback allowed us to examine which feedback metrics were useful and which ones led to better performance. In other words, while this study maintains that

the brief peer discussion in Peer Instruction is valuable, the focus is on settings in which such a discussion session is not efficient or possible.

Finally, it is important to clarify that students did not simply answer the questions for the second time, but they revised them, meaning that the students were able to see their initial answers and their initial level of confidence, along with the respective classroom feedback. This ensured that students did not have to rely on their memory, especially since each quiz included several questions.

Based on the above, this study focused on the following research questions:

- RQ1: How valid (i.e., accurate) and useful (i.e., indicative of the correct answer) is each feedback metric (i.e., percentage, confidence, preparation, past performance) during the revision (revote) phase of ARS activities?
- RQ2: What is the impact of each feedback metric (ibid.) on students' attitudes and performance during ARS activities?

3. Method

3.1 Participants and Domain

The "Business Development with Information Systems" course is typically offered in the third semester in the Department of Management and lasts 14 weeks (10 weeks of lectures, followed by 4 weeks of project activities). A total of 159 students volunteered to participate, while small monetary prices in the form of coupons were given at the end to the ten top-scoring students, as an additional incentive. The students were randomly organized by the system into four conditions according to the

feedback metrics they had access to: PERC (percentage), PERC_CON (percentage and confidence), PERC_PREP (percentage and preparation), ALL_1 (percentage, confidence, and preparation), PERC_PAST (percentage and past performance), ALL_2 (percentage, confidence, and past performance). Data analysis was based only on the 113 students that participated successfully in all phases of the study. The study had two rounds, each one focusing on a different set of feedback metrics. Table 1 presents the size and feedback metrics used in the two rounds of the study. During the second round of the study, the students of the PERC_PREP and ALL_1 groups were re-assigned to the PERC_PAST and ALL_2 treatment conditions, respectively, while the PERC and PERC_CONF groups did not change.

{Insert Table 1 here}

3.2 The SAGA Tool

The Self-Assessment/Group Awareness (SAGA) tool was designed and developed by the research team as a versatile web-based audience response system tool that could be used in a series of studies to analyze the potential of clickers on issues such as self-assessment, engagement, knowledge retention, group awareness, etc. In SAGA, students (a) denote their levels of preparation for the quiz (Figure 1), (b) provide initial answers to eight multiple-choice questions along with their levels of confidence (Figure 2), (c) receive feedback and revise their initial answers (Figure 3), and finally (d) discuss the correct answers in class with the instructor.

As mentioned earlier, SAGA offered feedback based on two objective and two subjective metrics (Table 2). Information on percentage and past performance was

calculated automatically by the system. On the other hand, information regarding students' levels of confidence and preparation was self-reported, and, therefore, subjective.

{Insert Table 2 here}

{Insert Figure 1 here}

{Insert Figure 2 here}

{Insert Figure 3 here}

3.3 Study Design

The study was conducted in two rounds, each applying a 4x1 between-groups experimental design as presented in Table 1. The independent variable was the feedback condition, while the dependent variables were the students' activity in the quiz (i.e., scores , revisions, levels of confidence and preparation, and past performance) and their responses in the activity evaluation questionnaire. During the second round, the preparation metric was replaced with the past performance metric in the respective groups. This was done because examining and comparing the past performance metric in relation to other metrics was possible only after a number of weeks had passed.

3.4 Procedure

In the beginning of the study, we informed students about the research nature of the course and the fact that they were going to be randomly assigned into different treatment conditions. We also ensured them that their participation would not affect their course grade, and we presented the monetary prizes for the top ten scores, urging them to stay engaged until the end.

The study started during the third week of the course and ended on the eighth, including two three-week long rounds. Lecture material was typically uploaded on the course's website a week in advance and although it was not mandatory, students were encouraged to study it before coming to class (this is common practice in Danish universities). As mentioned earlier, clicker activities can be used at different points during a lecture for different purposes. Focusing on identifying prior understandings and misconceptions, we used SAGA in the beginning of the course. Each week, students were starting the class by going through the four steps of the activity and were allowed 10 minutes for the first two steps (preparation and initial answers), five minutes for revisions, and five minutes to check their scores and the correct answers and discuss them with the instructor. After the 20 minutes, the week's lecture was starting, during which the instructor could revisit quiz questions and clarify misconceptions recorded in the results.

At the end of the study, students filled in a questionnaire, in which they stated their opinions about different elements of the activity. Students participated anonymously and individually in the activity, all being at the same step of the process at all times.

3.5 Data Collection and Analysis

The analysis of students' activity during the quiz (i.e., analysis of performance, revisions, confidence, preparation, etc.) was performed with the use of parametric tests. On the contrary, the analysis of students' responses in the activity questionnaire was performed with non-parametric tests, because in several questionnaire items the criterion for normal distribution of data was violated. For all statistical tests, a level of confidence at 0.05 was used, while statistical analysis was conducted with SPSS 20.

Analysis on students' activity was performed in two steps. First, we identified the questions that were more "challenging" for the students (i.e., questions in which students seemed divided between at least two choices during the initial phase) and then we analyzed students' activity in the two types of questions (challenging/non-challenging). This is similar to the third step in Peer Instruction approach presented earlier. In order to be consistent in evaluating the questions, we defined as challenging all the questions in which one of the two following statements were true:

- The percentage of students selecting the correct question choice was less than 50%.
- The percentage of students selecting the correct question choice was more than 50% and the difference between the correct and the second most favorite question choice was less than 20%.

These criteria were used to identify the challenging questions as one set (with two subsets, respective to the two study rounds). The argument for performing separate analysis on the challenging questions was that we expected students to rely significantly on the percentage metric, since they were more familiar with this concept. The need for the additional feedback was expected to have an impact in questions where students had diverging answers. Identifying those questions was also

interesting in analyzing student activity in general, since, as Wit (2003) argued, questions that yield divergent responses are more effective in stimulating discussion.

It was not feasible during the design of the study to predict which of the questions would be challenging for the students. Using the two conditions mentioned above, we identified 12 challenging questions in the first round of the activity and eight challenging questions in the second round. The analysis of the impact of the different feedback metrics was focused primarily on these two subsets of challenging questions.

4. Results

4.1 Metric Analysis

Table 3 presents a comparison of the two types of questions (challenging/non-challenging). According to our definition of challenging questions, the percentage metric provided a clear indication towards the correct answer in 28 questions, while it was unclear in the remaining 20 questions (i.e., 12 in the first and eight in the second round). T-test results revealed a significant difference in the percentage values of the correct choices between these two question types ($t[46] = 10.25$, $p < 0.01$, $d = 3.07$). Similarly, t-test results showed that the students performed significantly more revisions in the challenging questions ($t[46] = 4.31$, $p < 0.01$, $d = 1.29$), but with a significantly lower success rate ($t[46] = 4.73$, $p < 0.01$, $d = 1.41$).

{Insert Table 3 here}

Table 4 shows our analysis regarding the other three metrics among all questions and their ability to point to the correct choice. Paired-samples t-test results revealed significant differences between the average values of the correct and the second most selected choice for all metrics ($p < 0.01$). However, when considering the relative difference between those two choices, it appears that on average the confidence metric was 25.96% higher in the correct choice than in the second most selected choice, while the respective relative differences were 21.54% for the preparation and 4.74% for the past performance metrics. Relative difference is an important indicator of the usefulness of each metric, since it can be easily calculated by the student during the quiz activity, by examining the different values under each choice.

{Insert Table 4 here}

4.2 Activity Performance

Tables 5 and 6 present students' performance in the two rounds. One-way analysis of covariance (one-way ANCOVA) results revealed a significant effect of the treatment condition on the revised performance in the challenging subsets of both rounds, after controlling for initial performance (Round 1: $F(3, 108) = 4.64$, $p < 0.01$, $\eta^2 = 0.11$; Round 2: $F(3, 108) = 4.13$, $p < 0.01$, $\eta^2 = 0.10$). In the first round, post hoc tests revealed significant differences between the PERC group and the other three groups ($p < 0.01$, for all three), while in the second round, post hoc tests revealed that the PERC group was significantly outperformed by the PERC_CONF ($p < 0.01$) and the ALL_2 ($p < 0.01$) groups, but not by the PERC_PAST group that was comparable with all other groups ($p > 0.05$).

Finally, paired-samples t-test results showed that all treatment groups improved their performance significantly during the revision phase, in all cases ($p < 0.05$), with the exception of the PERC group whose initial and revised performance in the challenging subset of the first round were comparable ($p > 0.05$).

{Insert Table 5 here}

{Insert Table 6 here}

4.3 Student Activity

Repeated measures ANOVA results determined that the four treatment groups in both study rounds were comparable in terms of their self-reported levels of preparation, initial, and revised confidence ($p > 0.05$). In addition, paired-samples t-test results showed that students in all groups felt significantly more confident in the revision phase of the quiz ($p < 0.05$) (Table 7).

{Insert Table 7 here}

Pearson's bivariate correlation test results (Table 8) showed that preparation, initial confidence, and initial performance values were significantly correlated ($p < 0.05$), suggesting that students that felt prepared for the upcoming quiz, were also the ones that felt more confident during the quiz and achieved the highest scores, in the initial phase.

{Insert Table 8 here}

Regarding the number of revisions each group performed each week (Table 9), one-way ANOVA results showed no significant effect for the treatment condition ($p > 0.05$), with student performing similar amounts of revisions, regardless of the feedback metrics they had access to.

{Insert Table 9 here}

Question analysis showed that in nine challenging questions the most popular answer was not the correct and that its percentage increased in the revision phase. Moreover, statistical analysis on students that revised their answers showed that the PERC group had significantly less success in revising ($p < 0.05$) than groups that received the confidence and/or preparation metrics.

4.4 Student Opinions

Kruskal-Wallis test results showed no significant differences in students' opinions on the various aspects of the activity, as recorded in the activity questionnaire ($p > 0.05$) (Table 10).

{Insert Table 10 here}

Students said that the weekly quiz activities helped them prepare for the upcoming class (Q1a) and the final course examination (Q1b). Students' opinions on

the activity were even more positive regarding understanding the material and identifying errors and misconceptions (Q1c, Q1d). In evaluating the usefulness of the four feedback metrics, students in the respective treatment groups corroborated our expectation, identifying the percentage metric as the most useful one (Q2a). The second most useful metric was the confidence level (Q2b). Peers' past performance was evaluated as the third most useful metric (Q2d), while the preparation metric came fourth with students expressing mixed opinions (Q2c). We need to underline here that these findings depict metric usefulness as perceived by the students. Performance analysis earlier showed, for example, that groups that received feedback on preparation outperformed the PERC group, while students that received past performance information were comparable to it. Finally, students expressed an overwhelmingly strong opinion in favor of the activity saying that they would like its integration in additional courses (Q3).

5. Discussion

5.1 Usefulness and Validity of Feedback Metrics: Confidence and Preparation Outperformed Past Performance

The usefulness of each metric in guiding students towards the correct answer was affected by how valid and indicative the metric was. On one hand, preparation and confidence were self-reported and their validity was based on students' self-assessment. On that account, the bivariate correlations between preparation, confidence, and initial performance suggest that the students were accurate in self-assessing their levels or preparation and confidence. On the other hand, percentage

and past performance information was automatically calculated by SAGA. The percentage metric was able to point towards the correct choice in the majority of the questions, while metric analysis on the other three metrics (Table 4) indicated that it was easier for students to identify the correct choice by checking the values of the confidence and preparation metrics. Indeed, results showed that in the nine challenging questions in which the most popular answer was incorrect, students that received the additional feedback metrics chose not to follow the most populous answer group, but the most confident/prepared one. This suggests that students appreciated the metacognitive information they received on their peers. On the contrary, past performance scores were comparable in all questions, making this metric ambiguous. One explanation for the low impact of the past performance metric could be the fact that the quizzes, albeit belonging to the same course, were independent to each other, meaning that a high score in a week's quiz did not guarantee a high score next week.

Students' evaluation of the usefulness of four feedback metrics is consistent with our analysis regarding percentage and confidence, but diverges in relation to the preparation and past performance metrics. Students seemed to appreciate past performance more than preparation as feedback. Our assumption is that students may re-assess their level of preparation after seeing the quiz questions. This, however, means that they have a reason to mistrust this metric as inaccurate. Despite the fact that past performance metric has proven to be unhelpful, several students seemed to appreciate it. Our assumption is that the past performance metric is strongly linked to the notion of a strong/weak student, and therefore is very familiar to the students.

Analysis showed that in all cases confidence increased during the revision phase, following an increase in performance. Our results are in line with available literature

on the importance of asking students about their confidence during assessment (e.g., Kleitman & Costa, 2014; Schnaubert & Bodemer, 2015). At the same time, study findings suggest that although it may not be appreciated at the same level as confidence, information on peers' preparation may in fact be a useful feedback metric.

5.2 Attitudes and Performance: Positive Reception of the Tool and Learning Gains from Confidence and Preparation Metrics

Students expressed a strongly positive attitude towards SAGA, stating that the quizzes helped them prepare for the lectures and the final examination and that they were useful for them in better understanding the material and identifying errors and misconceptions. These learning benefits are cited often in the ARS literature (e.g., Mayer et al., 2009; Shapiro & Gordon, 2012, 2013; Shapiro et al., 2017, and Hoekstra & Mollborn, 2012, respectively) and refer to the activity as whole, appearing at the same level in all four treatment conditions. The latter suggests that even in their simpler feedback condition (PERC group), SAGA triggered strongly positive attitudes to students. However, this also hints that students' perceptions on clickers do not always agree with the actual learning benefits that the clickers support (the previously analyzed students' preference of past performance over preparation was also an example of this).

Regarding students' performance in the activity, results showed that the groups were comparable in the non-challenging questions. However, significant differences were revealed when the analysis focused only on the subsets of questions of the two rounds that we deemed challenging. Metric analysis mirrors exactly the findings of student performance in the two rounds of the study. Students that received only

information on the percentage of peers in each question choice had difficulties in the challenging questions, while students that additionally received the confidence and/or the preparation information (i.e. PERC_CONF, PERC_PREP, ALL_1, ALL_2 groups) scored significantly higher. Students that received information on percentage and past performance (i.e., PERC_PAST group) scored slightly higher than the PERC and lower than the other groups, with none of these differences reaching significance. These are clear indications on how information such as the level of confidence and preparation can be used as valid and useful feedback in ARS activities – especially in settings where direct peer interaction is not feasible. While lacking the dynamic nature of peer discussion, such metrics could still present valuable information on a large audience in a short amount of time. We therefore argue that their integration in audience response systems could offer additional learning gains.

5.3 Limitations

A certain limitation of this study is that the potential of the feedback metrics we tested was focused primarily on the subsets of questions we characterized challenging. One may argue that a different definition could lead to different subsets. To address this concern, we analyzed the validity of the question dichotomy in terms of percentage of the correct choice, the number of revisions made, and the success rate of revisions. Results suggest that the challenging questions posed, indeed, a more difficult task to the students.

6. Conclusions

The study provided empirical evidence on the beneficial role of using the level of preparation and confidence as feedback metrics in audience response systems. The metrics examined in this study provided a more detailed image of the classroom population and they were able to support students in their revision strategies at different degrees. In order to be useful, any metric, accompanying the percentage, needs to be both valid and indicative. In the current study, information on past performance was valid, but not indicative, while confidence and preparation information were effective, because students' self-assessment was sufficiently accurate. This may not be the case in a different setting, e.g., younger students, in which participants' metacognitive levels are lower.

Using a set of quantitative and qualitative variables, the study corroborated the findings reported in the literature on the positive effect the classroom response systems may have on student attitudes and performance. Moreover, the study pointed out how to effectively deal with misconceptions. This is a known strength of audience response systems, since they can provide timely feedback on individual and classroom level to students and instructors alike.

References

Anderson, L. S., Healy, A. F., Kole, J. A., & Bourne, L. E., Jr. (2013). The clicker technique: cultivating efficient teaching and successful learning. Applied Cognitive Psychology, 27(2), 222-234.

Barr, M. L. (2017). Encouraging college student active engagement in learning: Student response methods and anonymity. Journal of Computer Assisted Learning, 33(6), 621-632.

Blackwell, C. K., Lauricella, A. R., Wartella, E., Robb, M., & Schomburg, R. (2013). Adoption and use of technology in early education: The interplay of extrinsic barriers and teacher attitudes. Computers & Education, 69, 310-319.

Blasco-Arcas, L., Buil, I., Hernández-Ortega, B., & Sese, F. S. (2013). Using clickers in class. The role of interactivity, active collaborative learning and engagement in learning performance. Computers & Education, 62, 102-110.

Bojinova, E., & Oigara, J. (2013). Teaching and learning with clickers in higher education. International Journal of Teaching and Learning in Higher Education, 25(2), 154-165.

Boud, D., & Molloy, E. (2013). Rethinking Models of Feedback for Learning: The Challenge of Design. Assessment & Evaluation in Higher Education, 38(6), 698–712.

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.

Brooks, B.B.J, & Koretsky, M.D.M. (2011). The influence of group discussion on students' responses and confidence during peer instruction. Journal of Chemical Education, 88, 1477–1484.

Caldwell, J. E. (2007). Clickers in the large classroom: current research and best-practice tips. Life Sciences Education, 6(1), 9-20.

Chen, J. C., Whittinghill, D. C., & Kadlowec, J. A. (2010). Classes that click: Fast, rich feedback to enhance students' learning and satisfaction. Journal of Engineering Education, 99(2), 158-169.

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. Reviews in Physics Education Research, 1(1), 40-95.

Crouch, C., & Mazur, E. (2001). Peer Instruction: Ten years of experience and results. American Journal of Physics, 69, 970-977.

Dehler, J., Bodemer, D., Buder, J., & Hesse, F. W. (2011). Guiding knowledge communication in CSCL via group knowledge awareness. Computers in Human Behavior, 27(3), 1068–1078.

Desrochers, M. N., & Shelnutt, J. M. (2012). Effect of answer format and review method on college students learning. Computers and Education, 59, 946-951.

Drabman, R. S., & Lahey, B. B. (1974). Feedback in classroom behavior modification: effects on the target and her classmates. Journal of Applied Behavior Analysis, 7(4), 591–598.

Dufresne, R. J., Gerace, W. J., Leonard, W. J., Mestre, J. P., & Wenk, L. (1996). A classroom communication system for active learning. Journal of Computing in Higher Education, 7(2), 3-47.

Elicker, J., & McConnell, N. (2011). Interactive learning in the classroom: Is student response method related to performance? Teaching of Psychology, 38(3), 147-150.

Engelmann, T., Dehler, J., Bodemer, D., & Buder, J. (2009). Knowledge awareness in CSCL: A psychological perspective. Computers in Human Behavior, 25, 949–960.

Erkens, M., & Bodemer, D. (2015). Potentials of Text Mining based Technologies for Improving Cognitive Group Awareness. 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). Gothenburg, Sweden.

Ertmer, P. A., Ottenbreit-Leftwich, A. T., Sadik, O., Sendurur, E., & Sendurur, P. (2012). Teachers beliefs and technology integration practices: a critical relationship. Computers and Education, 59(2), 423-435.

Fies, C., & Marshall, J. (2006). Classroom response systems: a review of the literature. Journal of Science Education and Technology, 15(1), 100-109.

Fortner-Wood, C., Armistead, L., Marchand, A., & Morris, F. B. (2013). The effects of student response systems on student learning and attitudes in undergraduate psychology courses. Teaching of Psychology, 40(1), 26-30.

Gehlen-Baum, V., Weinberger, A., Pohl, A., & Bry, F. (2012). Technology Use in Lectures to Enhance Students‘ Attention. In Proceedings of EC-TEL 2012, LNCS 7563 (pp. 125–137). Berlin: Springer.

Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. Learning and Teaching in Higher Education, 1(8), 3-31.

Griffiths, L., & Higham, P. A. (2018). Beyond hypercorrection: remembering corrective feedback for low-confidence errors. Memory, 26(2), 201-218.

Han, J. H. (2014). Closing the missing links and opening the relationships among the factors: a literature review on the use of clicker technology using the 3P model. Educational Technology and Society, 17(4), 150-168.

Hancock, T. E., Stock, W. A., & Kulhavy, R. W. (1992). Predicting feedback effects from response-certitude estimates. Bulletin of the Psychonomic Society, 30(2), 173-176.

Hattie, J. (2009). Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement. London: Routledge.

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

Hoekstra, A. & Mollborn, S. (2012). How clicker use facilitates existing pedagogical practices in higher education: data from interdisciplinary research on student response systems. Learning, Media and Technology, 37(3), 303-320.

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, 315-325.

Kay, R. H., & LeSage, A. (2009). Examining the benefits and challenges of using audience response systems: a review of the literature. Computer & Education, 53(3), 819-827.

Kennedy, G. E., & Cutts, Q. I. (2005). The association between students' use of electronic voting systems and their learning outcomes. Journal of Computer Assisted Learning, 21(4), 260-268.

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.

Koenig, K. (2010). Building acceptance for pedagogical reform through wide-scale implementation of clickers. Journal of College Science Teaching, 39(3), 46-50.

Kulhavy, R.W., & Stock, W.A. (1989). Feedback in written instruction: the place of response certitude. Educational Psychology Review 1 (4), 279-308.

MacGeorge, E., Homan, E., Dunning, S., Elmore, D., Bodie, D., Evans, G., et al. (2008). Student evaluation of audience response technology in large lecture classes. Educational Technology Research and Development, 56(2), 125-145.

Marshall, L., Valdosta, G., & Varnon, A. (2012). An empirical investigation of clicker technology in financial accounting principles. Learning in Higher Education, 8(1), 7-17.

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.

McDonough, K., & Foote, J. (2015). The impact of individual and shared clicker use on students' collaborative learning. Computers & Education, 86, 236-249.

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.

Mollborn, S., & Hoekstra, A. (2010). “A meeting of minds”: Using clickers for critical thinking and discussion in large sociology classes. Teaching Sociology, 38, 18-27.

Mun, W. K., Hew, K. F., & Cheung, W. S. (2009). The impact of the use of response pad system on the learning of secondary school physics concepts: a Singapore quasi-experiment study. British Journal of Educational Technology, 40(5), 848-860.

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.

Nielsen, K.L., Hansen-Nygård, G., & Stav, J.B. (2012). Investigating peer instruction: how the initial voting session affects students' experiences of group discussion. ISRN Education, 2012, article id 290157.

Perez, K.E., Strauss, E.A., Downey, N., Galbraith, A., Jeanne, R., Cooper, S., & Madison, W. (2010). Does displaying the class results affect student discussion during peer instruction? CBE Life Science Education, 9, 133–140.

Poirier, C. R., & Feldman, R. S. (2007). Promoting active learning using individual response technology in large introductory psychology classes. Teaching of Psychology, 34(3), 194-196.

Prince, M. (2004). Does Active Learning Work? A review of the Research. Journal of Engineering Education, 93 (3), 223-231.

Sangin, M., Molinari, G., Nüssli, M.-A., & Dillenbourg, P. (2011). Facilitating peer knowledge modeling: Effects of a knowledge awareness tool on collaborative learning outcomes and processes. Computers in Human Behavior, 27(3), 1059–1067.

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.

Shapiro, A. M. (2009). An empirical study of personal response technology for improving attendance and learning in a large class. Journal of the Scholarship of Teaching and Learning, 9(1), 13-26.

Shapiro, A. M., & Gordon, L. (2013). Classroom clickers offer more than repetition: Converging evidence for the testing effect and confirmatory feedback in clicker-assisted learning. Journal of Teaching and Learning with Technology, 2(1), 15-30.

Shapiro, A. M., & Gordon, L. T. (2012). A controlled study of clicker-assisted memory enhancement in college classrooms. Applied Cognitive Psychology, 26, 635-643.

Shapiro, A. M., Sims-Knight, J., O'Rielly, G. V., Capaldo, P., Pedlow, T., Gordon, L., & Monteiro, K. (2017). Clickers can promote fact retention but impede conceptual understanding: The effect of the interaction between clicker use and pedagogy on learning. Computers & Education, 111, 44-59.

Siau, K., Sheng, H., & Nah, F. F. (2006). Use of a classroom response system to enhance classroom interactivity. IEEE Transactions on Education, 49, 398-403.

Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., Su, T. T. (2009). Why Peer Discussion Improves Student Performance on In-Class Concept Questions. Science, 122-124.

Stowell, J., Oldham, T., & Bennett, D. (2010). Using student response systems ("clickers") to combat conformity and shyness. Teaching of Psychology, 37(2), 135-140.

Strasser, N. (2010). Who wants to pass math? Using clickers in calculus. Journal of College Teaching & Learning, 7(3), 49-52.

Sutherlin, A. L., Sutherlin, G. R., & Akpanudo, U. (2013). The effect of clickers in university science courses. Journal of Science Education and Technology, 22(5), 651-666.

Vickrey, T, Rosploch, K., Rahmanian, R., Pilarz, M., & Stains, M. (2015) Research-based implementation of Peer Instruction: A literature review, CBE Life Sciences Education, 14(1).

Watkins, J., & Mazur, E. (2010). Just-in-Time teaching and peer instruction. In S. Simkins, & M. H. Maier (Eds.), Just in time teaching: Across the disciplines, and across the academy (pp. 39-62). Sterling, VA: Stylus Publishing.

Wit, E. (2003). Who wants to be... the use of a personal response system in statistics teaching. MSOR Connections, 3(2), 14-20.

Fig. 1 SAGA screenshot during the preparation question

33x8mm (300 x 300 DPI)

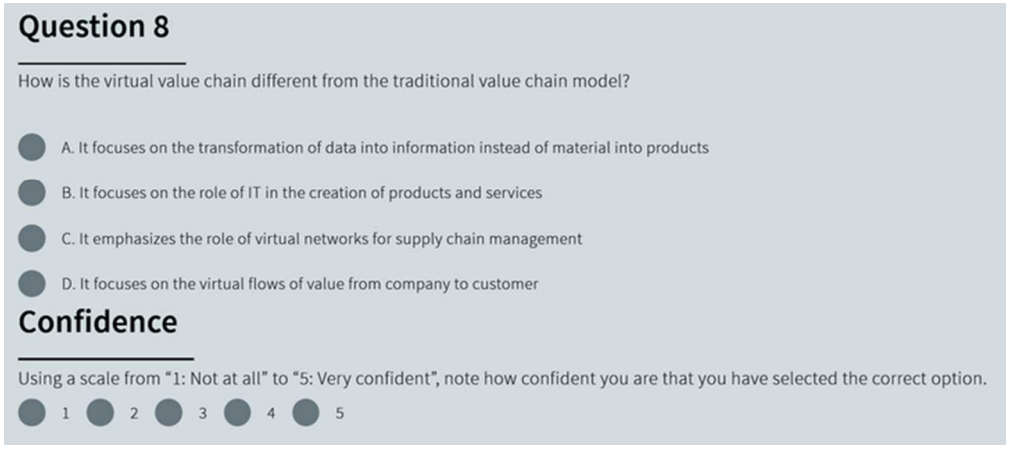

Fig. 2 SAGA screenshot during initial answer phase

61x26mm (300 x 300 DPI)

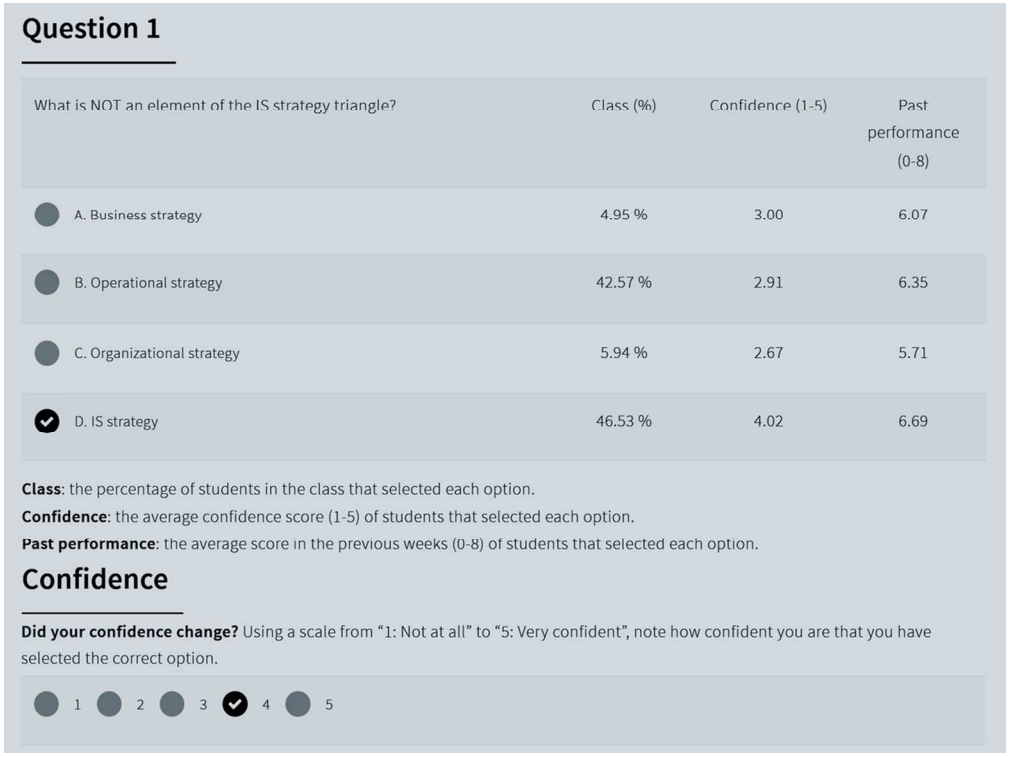

Fig. 3 SAGA screenshot during revision phase (ALL_2 group)

113x84mm (300 x 300 DPI)

Table 1. Treatment groups in the two rounds of the study.

FIRST ROUND				SECOND ROUND		
Group	**Feedback Metrics**	**Size**		**Group**	**Feedback Metrics**	**Size**
PERC	Percentage	28	–	PERC	Percentage	28
PERC_CONF	percentage and confidence	33	–	PERC_CONF	percentage and confidence	33
PERC_PREP	percentage and preparation	25	→	PERC_PAST	percentage and past performance	25
ALL_1	percentage, confidence, and preparation	27	→	ALL_2	percentage, confidence, and past performance	27

Table 2. Feedback metrics used in the revision phase of the study.

Metric (obj./subj.)	**Description** (scale)
percentage (obj.)	The percentage of the students in the class that selected each choice. (0%-100%)
past performance (obj.)	The mean score in all previous weeks of students that selected each choice. (0-8)
confidence (subj.)	The mean confidence score of students that selected each choice. (1-5)
preparation (subj.)	The mean preparation score of students that selected each choice. (1-5)

Table 3. Comparison between challenging and non-challenging questions.

	Challenging (n=20)		**Non-challenging** (n=28)	
	M	**SD**	**M**	**SD**
Percentage (%) of the correct choice during initial phase	39.96	(11.27)	75.72	(11.93)
Percentage (%) of students that revised their initial answers	35.08	(9.87)	22.40	(9.57)
Success rate (%) of correct revisions	60.63	(14.90)	80.44	(14.15)

Table 4. Metric analysis.

	Confidence			Preparation			Past Performance		
	M	**SD**	**% diff.†**	**M**	**SD**	**% diff.**	**M**	**SD**	**% diff.**
a. Correct choice	3.15	(0.53)		2.99	(0.22)		6.35	(0.36)	
			25.96%			21.54%			4.74%
b. Second most selected choice	2.50	(0.42)		2.49	(0.26)		6.07	(0.40)	

†. Relative difference (a-b)/b.

Table 5. Student performance in the First Round.

	PERC			PERC_CONF			PERC_PREP			ALL_1			
	M	**SD**	**n**	**M**	**SD**	**n**	**M**	**SD**	**n**	**M**	**SD**	**n**	**ANCOVA**
Non-Challenging (scale: 0-12)													
Initial	8.73	(1.76)	28	8.84	(1.51)	33	9.41	(1.61)	25	9.13	(1.68)	27	$p > 0.05$
Revised	10.74	(1.11)	28	10.75	(1.32)	33	11.16	(1.26)	25	10.58	(1.37)	27	
Challenging (scale: 0-12)													
Initial	4.19	(3.41)	28	3.50	(3.47)	33	4.27	(3.25)	25	4.12	(3.47)	27	$F(3, 108) = 4.64,$
Revised	4.30	(3.54)	28	6.08	(3.86)	33	6.39	(3.38)	25	6.84	(3.04)	27	$p = < 0.01, \eta^2 = 0.11$

Table 6. Student performance in the Second Round.

	PERC			PERC_CONF			PERC_PAST			ALL_2			
	M	**SD**	**n**	**M**	**SD**	**n**	**M**	**SD**	**n**	**M**	**SD**	**n**	**ANCOVA**
Non-Challenging (scale: 0-16)													
Initial	13.73	(1.50)	28	13.38	(1.81)	33	15.12	(1.64)	25	14.09	(1.67)	27	$p > 0.05$
Revised	16.63	(0.62)	28	16.66	(1.24)	33	16.63	(0.55)	25	17.22	(0.90)	27	
Challenging (scale: 0-8)													
Initial	2.10	(1.91)	28	2.52	(2.21)	33	2.80	(1.91)	25	2.34	(1.94)	27	$F(3, 108) = 4.13,$
Revised	3.41	(1.86)	28	4.78	(1.76)	33	4.23	(1.56)	25	4.57	(1.17)	27	$p = < 0.01, \eta^2 = 0.10$

Table 7. Students' perceived levels of confidence.

	PERC			PERC_CONF			PERC_PREP			ALL_1		
	M	SD	n	M	SD	n	M	SD	n	M	SD	n
Week 1												
Initial	2.92	(0.89)	28	2.86	(0.79)	33	2.83	(0.98)	25	2.85	(0.93)	27
Revised	3.26	(0.83)	28	3.09	(0.88)	33	2.94	(0.99)	25	2.93	(1.18)	27
Week 2												
Initial	3.20	(0.88)	28	2.95	(0.93)	33	3.00	(1.12)	25	2.86	(0.99)	27
Revised	3.73	(0.92)	28	3.37	(0.99)	33	3.43	(1.25)	25	3.24	(1.20)	27
Week 3												
Initial	3.12	(1.32)	28	3.12	(0.93)	33	3.21	(0.99)	25	2.85	(0.85)	27
Revised	3.50	(1.25)	28	3.55	(0.88)	33	3.41	(1.05)	25	3.23	(1.11)	27
	PERC			**PERC_CONF**			**PERC_PAST**			**ALL_2**		
	M	SD	n	M	SD	n	M	SD	n	M	SD	n
Week 4												
Initial	3.57	(0.96)	28	3.38	(1.19)	33	3.60	(1.16)	25	3.47	(1.34)	27
Revised	4.13	(0.82)	28	3.31	(1.79)	33	3.99	(1.21)	25	3.80	(1.39)	27
Week 5												
Initial	2.66	(1.00)	28	2.97	(1.08)	33	3.11	(1.27)	25	2.51	(1.31)	27
Revised	3.46	(1.00)	28	3.30	(1.16)	33	3.38	(1.48)	25	2.98	(1.49)	27
Week 6												
Initial	3.39	(1.26)	28	3.63	(0.96)	33	3.73	(0.93)	25	3.30	(1.27)	27
Revised	3.88	(1.18)	28	3.73	(1.31)	33	3.91	(1.08)	25	3.74	(1.27)	27

Table 8. Bivariate correlations.

		Preparation	Initial Confidence	Initial Performance
Preparation†	*r*	1	0.51	0.59
	p	-	0.00	0.00
	n	113	113	113
Initial Confidence†	*r*		1	0.58
	p		-	0.00
	n		113	113
Initial Performance†	*r*			1
	p			-
	n			113

†. Total average value.

Table 9. Revisions per group.

	PERC			PERC_CONF			PERC_PREP			ALL_1		
	M	SD	n	M	SD	n	M	SD	n	M	SD	n
Week 1	1.68	(1.52)	28	2.08	(1.64)	33	1.54	(1.77)	25	1.65	(1.67)	27
Week 2	2.05	(1.43)	28	1.18	(1.29)	33	2.13	(2.09)	25	1.56	(1.88)	27
Week 3	1.91	(1.50)	28	1.96	(1.79)	33	1.67	(1.62)	25	2.27	(1.95)	27
	PERC			**PERC_CONF**			**PERC_PAST**			**ALL_2**		
	M	SD	n	M	SD	n	M	SD	n	M	SD	n
Week 4	2.81	(1.47)	28	2.73	(1.45)	33	2.57	(1.93)	25	2.78	(1.62)	27
Week 5	2.00	(1.55)	28	2.17	(1.73)	33	1.78	(1.91)	25	2.18	(2.05)	27
Week 6	2.05	(1.77)	28	1.79	(1.75)	33	2.05	(1.93)	25	2.13	(2.17)	27

Table 10. Students' responses in the activity questionnaire.

Question			M	SD	n
Q1.	Has the activity helped you in … (1:No; 5:Yes)				
	a.	… preparing for the lesson each week?	3.62	(1.11)	113
	b.	… preparing for the final course exam?	3.64	(0.91)	113
	c.	… better understanding the course material?	3.97	(0.93)	113
	d.	… identifying errors and misconceptions?	4.12	(0.86)	113
Q2.	How useful was the … for you in revising your answers? (1:Not at all; 5:Very much)				
	a.	… percentage of students in the class that selected each option	3.96	(1.04)	113
	b. †	… average confidence score of students that selected each option	3.54	(1.18)	60
	c. ††	… average preparation score of students that selected each option	2.94	(1.37)	52
	d. †††	… average past performance of students that selected each option	3.34	(1.32)	52
Q3.	Would you like to have similar quiz activities in other courses?		Yes: 106; No:7		

†. Question available only to the PERC_CONF and ALL_1/ALL_2 groups.

††. Question available only to the PERC_PREP and ALL_1/ALL_2 groups.

†††. Question available only to the PERC_PAST and ALL_1/ALL_2 groups.