Ilkogretim Online - Elementary Education Online, 2020; Vol 19 (Issue 4): pp. 6270-6307
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2020.04.765033

Role Of Regular Descriptive Feedback In Enhancing Metacognitive Skills And Intrinsic Motivation: Evidence From Randomized Field Experiment

Dr. Muhammad Tariq Bhatti Assistant Professor, Department of Teacher Education, Faculty of Education, Shah Abdul Latif University, Khairpur, Sindh, Pakistan, mtbhatti2004@yahoo.com, tariq.bhatti@salu.edu.pk

Dr. Roshan Ali Teevno Assistant Professor, Department of Education, Begum Nusrat Bhutto Women University, Sukkur, Pakistan, roshan.teevno@bnbwu.edu.pk

Dr. Shila Devi Assistant Professor, Sindh Teacher Education Development Authority Karachi, Sindh, Pakistan, shiladevi@gmail.com

ABSTRACT

This study uses a hybrid approach of audience response systems (ARSs; such as Socrative) and revised Bloom's taxonomy (RBT) to examine how regular descriptive feedback affects students' metacognitive skills and intrinsic motivation. Through randomized field experiment (RFE), we divide 966 physics students from 16 colleges into three groups: control, individual, and cooperative group. The results show that cooperative treatment improves students' intrinsic motivation (IM) and metacognitive skills (MS). Individual treatment (InT) had no effect on IM or MS. Furthermore, mediation analysis suggests that IM mediates the CoT and MS. Based on these findings, we recommend combining descriptive feedback and peer discussion in physics classes with ARS technology assessment.

Keywords: Peer descriptive feedback (PDF), Audience response systems (ARSs), Randomized field experiment (RFE), intrinsic motivation, Metacognitive skills, Socrative, Regular Descriptive feedback (RDF).

1. **INTRODUCTION** Students are encouraged to learn and grow by interacting with their professors and classmates, when they receive descriptive feedback (DF) from their professors or peers (Armstrong et al., 2007; Bruff, 2009; Cain & Robinson, 2008). Specifically,

descriptive feedback (DF) allows students to clear up misconceptions and reveal gray areas in knowledge and skills. Regardless of the importance of metacognitive skills (MS), students rarely receive the required DF in traditional classroom environment, because the professor is unable to provide DF due to overcrowded classrooms, limited availability of resources and time constraints. Since MS are critical in the learning system (Mayer, 2008; Mayer et al., 2009; Petto, 2019), because research shows that intrinsic motivation (IM) and MS are strong predictors of learning and learning outcomes. As a result, professors must be aware of their students' current level of learning and growing learning needs to provide the DF in a strategy-focused program within constraints of time, resources, and congested classrooms. Previous metacognitive-focused literature indicates that answering the higher-order questions (HOQs) by using audience response systems (e.g., Kahoot, ARS, Socrative, and/or clickers) is effective and creates opportunities for providing immediate DF that meets the demands of meaningful learning (Mayhew et al., 2020; Temitayo; 2020). Students while responding these higher order questions (HOQs), about the learning content being applied, they receive DF and can use MS, if the content is meaningfully internalized and generally understood. It promotes to learning in two-ways, firstly, by encouraging students to discuss course content in the class, secondly, it gives professors insight into the students' thinking and mastery of content (Williams et al., 2016).

According to Mayhew et al. (2020), Paniagua and Swygert (2016), and Pearson (2019), metacognitive control and monitoring are achieved by using, this type of questioning (HOQs) in conjunction with regular descriptive feedback (RDF); assessing how well students have planned their approach to a learning task, use appropriate skills, and strategies to solve problems, monitoring one’s own comprehension of text, self-assessing, and self-correcting, evaluating progress towards the completion of task, and overall performance. As higher-order questions (HOQs) are presented with course content (e.g., meaningful concepts and themes) in situations that encourage students, by using more MS, they build up an extensive toolkit of metacognitive self-control and monitoring activities. What we are seeing here is consistent with research showing that students who receive regular descriptive feedback (RDF) are more motivated in the long run (Abrahamson, 2006; Åsta et al., 2018; Matthew & Ziemann, 2020; Mayhew et al., 2020; Perera & Hervás-Gómez, 2021) and intrinsic motivation (IM) to get new MS (Butler, 2018; Petto, 2019; Perera & Hervás-Gómez, 2021; Schell & Butler, 2018).

ARSs technology research has recently focused primarily on the relationship between formative assessments and their effects (Abdullah, 2018; Abrahamson, 2006; Åsta et al., 2018). Despite the fact that this research sheds light on important issues, there is a dearth of empirical evidence about why student performance on average improves. We draw the conclusion from previous research that more empirical evidence is needed to show how those students' MS are influenced by regular descriptive feedback (RDF) during formative assessment, using audience response systems (ARSs) structures (Abrahamson, 2006; Paniagua & Swygert, 2016; Sajna & Premachandran, 2016). Some previous research shows that students assess their learning status entirely on the form of available prompts during formative assessment (Temitayo, 2020; Williams et al., 2016), for example. The impression of familiarity in authenticating solution strategies and the response rate to a query are examples of indicative prompts. Well-designed strategy-focused treatments with metacognitive activities are needed to guide students in identifying indicative prompts. These activities improve the use of diagnostic prompts that predict future learning, academic achievement, and overall performance (Åsta et al., 2018; Vallely & Gibson, 2018; Yee et al., 2020). The most frequently used metacognitive cue during formative assessments with ARSs systems is regular descriptive feedback (RDF), which includes professor feedback and peer discussion. RDF, which refers to students overwriting, conformity, tune or restructuring of his memory, irrespective of whether or not this data is of knowledge domain, cognitive approaches, or metacognitive information (Åsta et al., 2018; Suzan et al., 2020).

1.1. Previous Studies

A few quasi-experimental studies explored how students develop their MS, in the classrooms that use audience response systems (ARSs; e.g., Socrative). According to Petto (2019), when students are assessed with formative assessment and audience response systems (ARSs, e.g., Socrative) (Barchilon Ben & Ben-Av, 2016), their MS improve. By using ARSs in the classroom, students receive several opportunities to obtain instant descriptive feedback (DF) and to test their understanding of the learned material. RDF refers as a medium and is a predictor of students' cognition development. ARSs tool use has a high-quality impact on the teaching-learning process, according to Yee and Ean (2020), who also found that MS and overall performance improved when ARSs technology use was combined with pedagogical techniques and RDF. Despite this, the study by Yee and Ean (2020) lacked a control group and a randomized research approach. Furthermore, this study was conducted at the university level with a limited number of groups of students. We used a randomized filed experiment (RFE), in this current study to assess the effects of professor's RDF and peer's descriptive feedback (PDF) (peers discussion), while using ARSs technology (Socrative) on students' metacognitive skills (MS) and intrinsic motivation (IM) in college physics education. Every phase students encouraged to answer conceptual higher order questions (HOQs; e.g., combination of the use of Socrative technology and higher-order questions based on the higher order thinking levels of the Revised Bloom's Taxonomy), facilitated by the internet ARSs (www.socrative.com) (Barchilon Ben & Ben-Av, 2016). Students in the cooperative and individual groups, in this current study, answer the HOQs individually. Students in the cooperative group (CoG) communicate their responses with peers before subject learning and discussion with professor, to assess the impact of professor's RDF with a combination of PDF and only professor RDF. Individual group (InG) students, on the other hand, do not share their responses with their peers and instead receive feedback directly from the professor. And students in the control group (CG) did not receive feedback from either the professor or their peers, but instead submitted their answers on an individual basis.

In daily teaching-learning routine, students are encouraged to monitor and regulate their personal learning (metacognitive skills) by receiving regular descriptive feedback (RDF) from peers and professors, which is the foundation of self-regulated learning. Student learning and outcomes can be improved as a result of this (Abrahamson, 2006; Åsta et al., 2018; Matthew & Ziemann, 2020; Pearson, 2019; Ward et al., 2017); it alters the students' intrinsic motivation (IM) and focus of attention at some stage of mastery learning, contributing to the satisfaction of the students' learning needs of feeling competent learner (Masania et al., 2018; Russell, 2020). According to Mountford-Zimdars et al. (2017) and Williams et al. (2016), RDF activated through higher order questions (HOQs) encourages students to experience the mastery of their learning and regulate the teaching-learning process. HOQs can be used in the classrooms, by applying the different techniques in multidimensional contexts. Professors can collect responses by asking students to indicate their response options by raising colored flashcards, applause, displaying small whiteboards, or raising their hands using simple daily routines or low-cost material (Flavell, 1976; Flavell, 1979; Flavell, 1987; Guarascio et al., 2017; Grahame, 2016; Suzan et al, 2020). Despite, these low-cost material are inexpensive, low-cost material have various drawbacks too. First of all, calculating the exact number of polling, particularly in large classrooms, is complicated for a professor in a limited period. In addition, students might feel stress of social compliance because of a fear of being exposed and may therefore be unable to respond honestly (Abdulla, 2018; Barchilon Ben & Ben-Av, 2016; Brady et al., 2013; Brady et al., 2015).

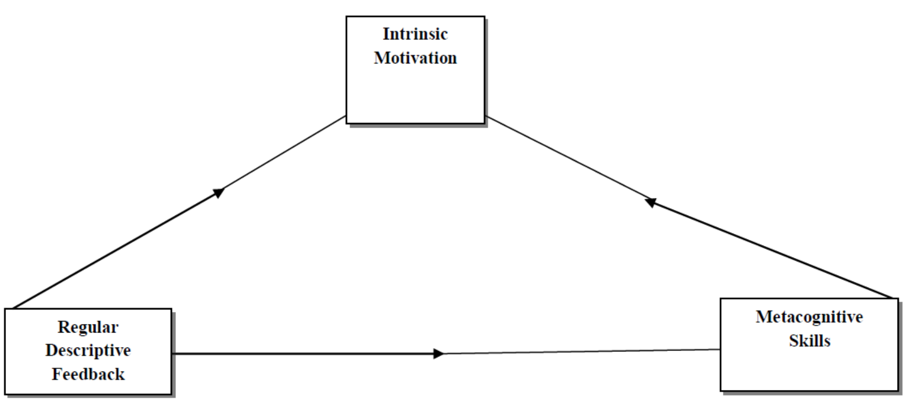

Fig. 1. Relational Model.

In a High Tech Assistive Technology (HTAT) setting, the technology (such as; ARSs- audience response systems) is used for the student response system in which students can enter their responses in a number of technology options, such as Kahoot (Butler, 2018; Cain & Robinson (2008), clicker technology (Bale, 2018; Compton & Allen, 2018), or the internet-based gadget Socrative (Cain & Robinson, 2008; Ruth & Shaista, 2020; Russell, 2020). This ARSs technology addresses the above limitations directly, fast collection, storage and display of a wide range of correct and incorrect answers on the front of the class screen, without any specific attribution. Because the effects are evaluated in a more private manner than feedback given in a low-cost setting, professor feedback based on technology is more appropriate and less intimidating to students' self-esteem (Vallely & Gibson, 2018; Wengrowicz et al., 2018). ARSs technology assessments provide students the opportunity to obtain immediate regular descriptive feedback (RDF) in a single class-session, multiple times, and in a variety of ways. Giving feedback is as simple as placing a check mark next to the correct answer as the professor reads it aloud. Students will only be told if they answered the question correctly; they will receive no other feedback. This limited amount of feedback, according to Zohar and Dori (2012), improves the achievement of students in comparison to conditions receiving no HOQs and no RDF throughout the lecture.

The tracking system gathers all class responses and provides feedback by displaying polling statistics of students' responses (for example, chart or graph), in front of the class with a tick mark next to the appropriate response, in a more effective and efficient way of ARSs technology use. Students can use polling data, to see how well they compare to their peers' knowledge, when used in this way. If their answers are incorrect, it helps students to realize they are not the only ones who struggle with the educational material, which boosts self-confidence and reduces the belief that they are incapable of understanding the educational material (Draper & Brown, 2004; Flavell, 1976; Flavell, 1979; Flavell, 1987; Hake, 1998). The graphic description of the students' responses provides professors with feedback so that they can explain the right answer and advise students about typical errors. Abdullah (2018), Astrid et al. (2020), and Hunsu et al. (2016) all concluded that displaying a variety of options and obtaining professor RDF, than just giving the right answer, has a significant impact on student learning and increases the impact of class-formative assessments and exams. Nevertheless, Bowen (2012) and Deslauriers et al. (2019), research asserted that professor RDF alone may be insufficient for the students. Students may find insight of content-focused terms problematic, and as a result, they indulge in difficulties when attempting to utilize that in their learning procedure. A more comprehensive approach to teaching RDF can make professor feedback more understandable for students when combined with peer descriptive feedback (PDF) (peer discussions). Peer discussions are descriptive by nature and are part of a cooperative learning strategy that evaluates learner's educational-content knowledge, in a more user-friendly language than just professor RDF or ICT alone (David et al., 2015; Heaslip et al., 2014; Howell et al., 2017; Veenman & Elshout, 1995; Zimmerman, 2000). Rather than using content-focused terminology, students with similar backgrounds can explain problems and provide answers in different ways than a professor (Coleman, 1998; Pearson, 2019; Reay et al., 2008; Ruth & Shaista, 2020; Suzan et al., 2020). When students interact with their co-learners, they can use their reasoning to arrive at an answer by explaining and supporting it (Preszler, 2009; Temitayo, 2020, Ward et al., 2017; Williams et al., 2016; Zimmerman, 2000).

As a common strategy and popular way to facilitate PDF, peer discussion involves having students first answer HOQs on their own, then discussing the reasoning behind their answers with their classmates, and finally answering the question once more before the final answer is shown and explained (Alexander, 2017; Astrid van et al., 2020; Badia et al., 2016; Paniagua & Swygert, 2016; Russell, 2020). This style of reasoning encourages students to recall, integrate, assimilate, regulate, and accommodate their existing learning with new learning. Armstrong et al. (2020), Deslauriers et al. (2019), Valiente et al. (2016), and Åsta Haukås et al. (2018), demonstrated that HOQs, is supported by PDF broaden metacognitive skills (MS) for determining knowledge, affected metacognitive monitoring too. RDF, for example, provided by professors or co-learners, on learner's understanding of content-focused material can spark intrinsic motivation (IM) (See Fig.1). Particularly, when RDF allows students to evaluate their learning outcomes and learning activities, it stimulates students to narrow the gap between what is learnt and what is aimed to be taught (Collins, 2007; Cronbach, 1951; Hung, 2016). Active learning and RDF through formative assessment (FA) with ARSs technology provide scaffolding for the development of students' IM, according to Baker et al. (2017) and Barchilon and Ben (2016).

Because of the ARSs technology, student-to-student and professor-to-student discussions become more proactive and focused, allowing professors to create dynamic and enjoyable lesson plans. ARSs technology offers students anonymity by keeping their answers private and not revealing them to their peers or professors right away. This is an appealing feature for students. It minimizes students' fear of failure (Astrid van et al., 2020; Collins, 2007; Crouch & Mazur, 2001; Hunsu et al., 2016), and increases students' active participation and involvement in the teaching-learning process (Brydges & Butler, 2012; Butler, 2018; Cameron & Bizo, 2019; Burton, 1991; Collins, 2007; Cronbach, 1951; Crouch & Mazur, 2001). More specifically, this anonymity may also encourage insecure and passive students to participate actively too, by electronically responding to HOQs and allowing them more time to process information, contributing energetically to class discussions (Chen et al., 2010;

Chou, 2017; De Corte et al., 2004; Dembo & Seli, 2007; Melanie & Christopher, 2018). Hence, a secure and safe learning atmosphere boosts the students' IM (Cubric & Jefferies, 2015). FA with ARSs technology use, for example, can only increase students' increased metacognitive awareness when interactive metacognitive activities such as these provide immediate RDF, inspire monitoring and increase students' IM (Chou, 2017; Hung, 2016; Crouch & Mazur, 2001; Masania et al., 2018; Mayhew et al., 2020).

Based on the previous findings, quizzing with ARSs technology with immediate RDF can boost the students' metacognitive skills (MS) and IM. Nevertheless, further empirical investigation is needed to determine: (i) the extent to which a cooperative (peer) treatment (CoT) and individual treatment (InT) improves MS and IM; and (ii) whether these strategies for RDF benefit some students more than others. Empirical evidence is available that there is a positive significant relationship between professors RDF or PDF and development of MS and IM at one side, and professor RDF or PDF and gender on the other side (Chi et al., 1994; Karpicke, 2012; Knight & Wood, 2005; Vallely & Gibson, 2018; Yee et al., 2020). Johns et al. (2012) asserted students who lack IM and have low MS benefit the most from co-learner discussions during FA when using ARSs technology. According to Chng and Gurvitch (2018), HOQs develop low-metacognitive aware students to the same extent as their more inspired and self-aware peers. However, the magnitude of the effect was lower in learners with high MS, indicating that the HOQs used required fewer problem-solving skills. Concerning gender and feedback, Chou (2017), Collins (2008), Pink (2009), and Little (2016) agreed that use of academic activities with ARSs technology no longer discriminate against gender and that it may be less complex for participation of any gender. Previous studies (Crossgrove & Curran, 2008; Crowe et al., 2008; Cotton, 2005; Mazur, 1997; Schell & Butler, 2018; Temitayo, 2020; Vallely & Gibson, 2018; Veenman & Elshout, 1995; Ward et al., 2017; Yee et al., 2020) found that interactive learning strategies benefit both boys and girls, but that they approach learning in different ways (Coleman, 1998; Matthew & Ziemann, 2020; Mayhew et al., 2020; Mayer, 2008; Mayer et al., 2009; Zimmerman, 2000). Boys are more competitive and less focused on RDF, whereas girls prefer learning while they may express themselves through peer interactions (Collins, 2008; Eva & Regehr, 2011; Forest, 2012; Mazur, 1997; Preszler, 2009; Zohar & Dori, 2012). The studies conducted by Masania et al. (2018), Collins (2008), and Chou (2017) found that while boys do better with MS in an individualized learning environment, girls do better when they use ARSs technology in a collaborative learning setting.

1.2. Conceptual framework

Based on the literature concerning IM, descriptive feedback (DF) during audience response systems (ARSs) technology use, and MS, in this paper's empirical section, we have developed a conceptual framework for evaluating the critical construct RDF, meanwhile, RDF may have an immediate impact on MS. Hence, RDF may have an indirect effect on MS through their influences on IM (Fig. 1). According to some metacognitive researchers, for example Guarascio et al. (2017) Heaslip et al. (2014) metacognitive skill (MS) is the ability to control one's cognitive processes and is highly correlated to mastery learning and outcomes (Howell et al., 2017; Little, 2016). It includes skills that allow students to consider, comprehend, and monitor their learning, it relates to higher-order thinking (Deslauriers et al., 2019). Learners with higher MS are better able to regulate and monitor their cognitive processes in a way that maximizes mastery learning and to succeed in achieving the learning goals (Cameron & Bizo, 2019; Herreid, 2010; Mshayisa, 2020; Pink, 2009). These learners define their learning goals, control, and monitor and regulate their behaviors, IM and MS in order to achieve these goals, self-directed learning (metacognition) which is a key construct in this study [30]. It has long been believed that IM helps students learn by keeping them interested in the task at hand. This results in the development of MS and mastery learning, as well as improved performance and results (Chng & Gurvitch, 2018; Mayhew et al., 2020; Vickrey et al., 2015). It also produces and regulates the ability to concentrate on a learning task and finish it, which is IM [88]. Lack of focus indicates that the intrinsic motivation to work on a learning task is lacking, and extrinsic motivation (EM), or the motivation to work on a task as a means to an end, is required to keep the focus on the learning task. A lack of focus indicates a lack of intrinsic motivation to work on a learning task (Chng & Gurvitch, 2018; Howell et al., 2017; Krause et al., 2017). This can happen if the learner's feedback shows a lack of ability, which hampers the motivation and following preference to carry out the learning task. While performing duties, intrinsic and extrinsic motivation are intertwined (Mayhew et al., 2020), in this current study we focus only on the intrinsic motivation (IM).

2. CURRENT STUDY

Our contribution to the literature is organized as follows: firstly, as far as our knowledge is concerned, the current study is one of the first empirical studies that investigate the impacts of professor regular descriptive feedback (RDF), employing ARSs technology (Socrative) on metacognitive skills (MS) and intrinsic motivation (IM), whether or not paired with peer descriptive feedback (PDF). With sufficient statistical rigor, we accomplish this by carrying out a randomized field experiment (RFE). Previous research has been quasi-experimental in nature, with pre-test and post-test designs but no control group (Barchilon Ben & Ben-Av, 2016; Cain & Robinson, 2008; Moh. Irma et al., 2021; Mountford-Zimdars et al., 2017; Vallely & Gibson, 2018; Yee et al., 2020). Secondly, we investigate the impacts in colleges, while previous research has concentrated on the higher education in the context of university. Moreover, research in this area, particularly in college education, is deemed critical, given that ARSs generation usage in college education is quickly increasing (Masania et al., 2018). Thirdly, we examine the effects of distinctive types of descriptive feedback (DF), on two treatment groups including a control group (CG), allowing us to assess differences in intrinsic motivation (IM) and metacognitive skills (MS). Furthermore, we can demonstrate various effects and varied impacts on heterogeneous populations using formative assessment (FA). For example, we can estimate effects and impacts for girls and boys individually, as well as across pre-treatment stages of the students' MS. We, fourthly, combine the ARSs technology (Socrative) with the revised Bloom's Taxonomy for generative higher order questions (HOQs, e.g., combination of the use of Socrative technology and higher-order questions based on the higher order thinking levels of the Revised Bloom's Taxonomy) and their implementation for formative assessment (FA), to investigate the effects of these questions on students' MS and IM. Finally, as we analyze each level of intrinsic motivation (IM) and MS, we can determine the mediating influence of IM on students' MS.

The implications of DF strategies (e.g., RDF--professor regular descriptive feedback and/or PDF--peer descriptive feedback) are analyzed for intermediate outcome intrinsic motivation (IM) and outcome metacognitive skills (MS). As a result, our research questions are:

1. What impacts do professor regular descriptive feedback (RDF) or peer descriptive feedback (peer discussion) (PDF) integrated with professor RDF have on metacognitive skills (MS)?
2. What impacts do professor regular descriptive feedback (RDF) or peer descriptive feedback (peer discussion) (PDF) integrated with professor RDF have on intrinsic motivation (IM)?
3. Is there a difference in the effects for different student subgroups?

a. Do girls and boys have different outcomes?

b. Does it make a difference, if the student has a low, moderate, or high MS, if the effects differ?

4. Is the impact of professor RDF or PDF integrated with professor RDF, on students' MS mediated (in part) by its impacts on IM?

In section two, a high-level review of the literature and a conceptual framework are presented, followed by the data and descriptive statistics, and the empirical design in section three. Section 4, discusses the findings and finally, section 5 brings the paper to a conclusion by discussing the results.

3. RESEARCH METHOD

3.1. Participants

More than 966 students from 16 colleges in Pakistan's southern areas took part in this large-scale treatment. All participating colleges are situated in Pakistan's heavily urbanized areas, and are typical colleges for K12 education. All data are within 0.5 SD (standard deviation) of the mean of all constructs for these specific colleges. The principal author can provide the representativeness analysis results upon request. Thirty five trained and experienced physics instructors took part in this treatment. Total of nine hundred and ninety-five physics students in fifty classes, K9 through K12 are the participants of this treatment. The pre-survey was completed by all 995 students from those participating colleges at the beginning of treatment, but twenty-nine students are excluded from the data analysis because they occasionally took part in weekly assessments (N =10), left class or college during treatment (N = 08), or did not complete the post-survey at the end of treatment (N = 11), and thus did not meet the treatment conditions sufficiently. Finally, 966 students included in the sample, networked within fourteen classes with the control condition (n=222), fourteen classes with individual condition (n=279), and twenty-two classes with cooperative condition (n=465). In addition, we use t-tests to compare the included (N = 966) and excluded (N = 29) students to see if they form a distinct group that shares at least one of the background characteristics with each other. In this study, no statistically significant variations in student characteristics between groups are validated.

3.2. Individual group

Individual group (IG) students used their digital devices to answer the same HOQs as the control group (CG) in the same time span, presenting the difficulty level of the HOQs in a comparable way across the treated and un-treated groups. Students in the control group can see the correct answer, but the results of HOQs for students in the individual group are anonymously displayed on a screen display in front of the class as a histogram with a distribution of answers. Students get a tick mark when they answer correctly, and their professor explains and informs them about the reasoning errors that underlie incorrect answers (Fig. 2).

3.3.Cooperative group

The format of this treatment is similar to that of Johns et al. (2012), here, students who use a cooperative group (CoG) feedback strategy (peer descriptive feedback combined with regular descriptive feedback from the professor) and those who use only an individual group strategy of feedback (professor regular descriptive feedback) answer a set of HOQs every week using ARSs polling technology. Unlike Johns et al. (2012), however, we employ a control group and administer a treatment to two treatment groups that can be compared to this control group. We also use pairs of questions of HOQs. These questions help students think about how they can influence their peers with the correct answer, rather than just copying it from their professors or peers. It gives students the opportunity to examine their problem-solving methods and develop reasoning and metacognitive skills (Badia et al., 2016; Baker et al., 2017; Bowen, 2012; Mayer, 2008; Mayer et al., 2009). Students learn via professor RDF or PDF, and as a result, they are instantly able to answer 2nd HOQ properly (Barber & Njus, 2007; Brady et al., 2013; Brady et al., 2015). For this treatment, we developed a database of 900 paired sets of adapted HOQs of consistent Pakistani National Curriculum of Physics (PNCP), examination conducted by the Boards of Intermediate and Secondary Education (BISE Sindh, BISE Punjab, and BISE Baluchistan). In each set of questions, the cognitive domain is assessed at the same level (Revised Bloom's Taxonomy—RBT, 2001) and conceptual understanding is assessed at the same level, especially at the analysis or evaluation level (Anderson & Krathwohl, 2001). This database then contains 565 pairs of software-level questions and 335 pairs of evaluation-level questions. The sets of HOQs include the entire core material of the Pakistani secondary and higher secondary physics program (e.g., electromagnetism, modern physics, waves, mechanics, and thermodynamics). A professor chooses five paired sets of HOQs (a total of ten questions) for students to solve using their digital devices and ARSs technology (Socrative), depending on the theme of the week (for example, see Fig A.1 in Appendix A).

The professor does not instantly provide the correct answer in the cooperative group after polling a HOQ, and the histogram of the students' polling is also not displayed. The HOQs are set to appear twice in this scenario, and they are accompanied by a peer educational technique (peer discussion) (Hake, 1998). Before answering the question again individually, students have four minutes to discuss their responses and alternative answers in pairs (think-pair-share). For the time being, the professor simply walks around the classroom, listening to and interacting with students as they discuss various topics (focusing on intrinsic motivation not just answers). A histogram of the previous responses is used to quickly locate the answers, and the students are then asked to answer the same question once more. The professor discusses the correct and incorrect answers after the tick mark verifies the correct response (Fig. 2). The second follow-up HOQs are given to students individually, and their responses are required, as they did in the control and individual groups, see dotted blocks in Fig. 2. A tick mark shows the students of the correct answer, after polling on this question. The professor provides students with immediate regular descriptive feedback (RDF) on how well the concepts are comprehended, after showing the histogram of the student responses. HOQs are asked five times per week in the manner depicted in Fig. 2.

3.4. Control group

To be effective, students in the control group (CG) must see a tick mark next to the correct answer on HOQs. They don't get any feedback on the polling data from their professor or peers, and they don't get any explanations for correct and incorrect responses either (Fig. 2). By providing all participants in this treatment with access to a secure website where they can read the restricted overall feedback of all correct and incorrect answers after class, we address the ethical consideration of having a control group that makes no benefit. We don't know, however, if the students made use of it on a regular basis.

3.5. Treatment

There were two treatment groups (a cooperative (peer) group and an individual group) as well as one control group (an untreated group) for which we devised an intervention using ARSs technology to evaluate our conceptual framework and provide answers to our research questions. Cooperative students use PDF (peer discussion) and professor RDF, while individual group students only get feedback from the professor. Every week for twenty-four weeks, students in all three groups (a control, a cooperative, and an individual groups) solved average 10 HOQs, which were comprised of five pairs of questions in each set. Questions with different numerical values and contexts validate the same conceptual understanding, but in different contexts, and with different concepts or principles (Krause et al., 2017). Researchers with prior experience using Revised Bloom's taxonomy (Anderson & Krathwohl, 2001; Armstrong et al., 2007) to investigate physics questions devised all weekly questions to test higher order cognitive skills on the 'analyzing' and 'evaluating' levels for higher order cognitive skills. Using 'analyzing' and 'evaluating' questions to allow students to demonstrate mastery of the learning martial improves students' learning ability and metacognitive development by studying and understanding the data by breaking it down (Anderson & Krathwohl, 2001; Armstrong et al., 2007; Beatty et al., 2006). There are four possible answers to each HOQ, with only one of them being correct. We use Socrative (ARSs), a free online student response software program that works on any digital device that can connect to the internet, because nearly all students have personal digital devices. Professors can easily manage formative assessments for their students with this online tool, which is accessible at www.socrative.com. Students can create problems and organize the flow of questions (Cameron & Bizo, 2019; Compton & Allen, 2018; Deslauriers et al., 2019). A computer or a laptop could be borrowed from the college by students who did not have their own digital devices or whose batteries ran out. This has happened a few times now. Meanwhile, the overall effects of using ARSs technology on digital devices as the primary device should have been unaffected. There is a large screen in front of the class and digital devices are used to show the questions.

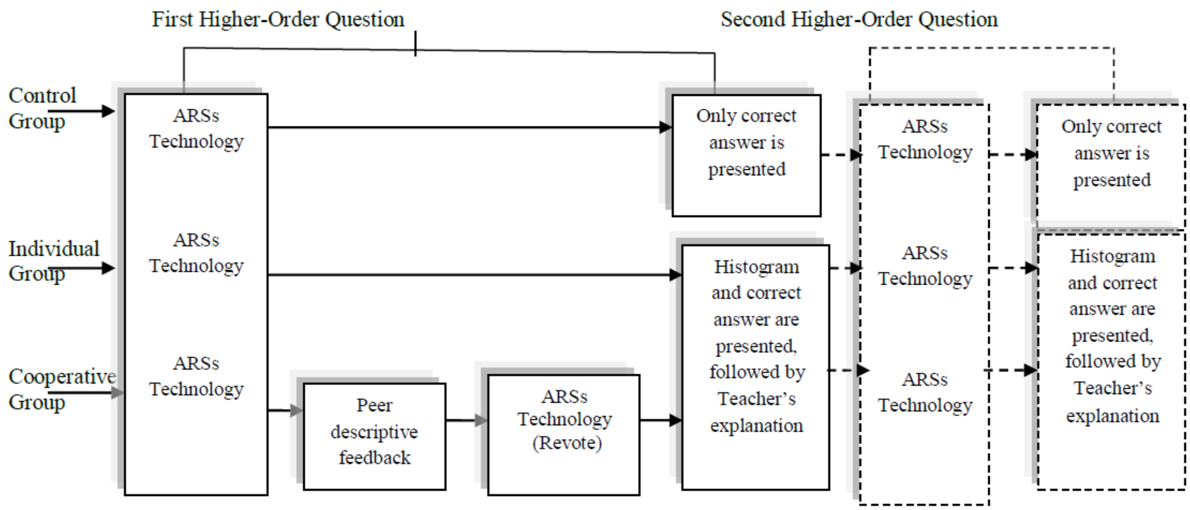

Fig. 2. Design of the Experiment.

Using their individual digital device and without communicating with one another, students are given three to four minutes to answer a question, depending on the difficulty and nature of the HOQs. The answers are then collected using the Socrative (ARSs technology) application, after the polling has been cast. The Socrative (ARSs) program records all student responses, including null responses. Because their responses aren't associated with their real names, students are anonymous in the eyes of the professor. It doesn't matter whether or not the responses are correct because they are not graded and have no bearing on the students' final grades. Even though all three groups apply the same Socrative application (ARSs technology), how feedback is presented differs among groups.

3.6. Treatment time detail

Sixteen nearby colleges were chosen, a few months before the treatment began, based on the researchers' previous contacts. This information was provided to them via e-mails from the principals of participating institutions, who then asked if they wanted to take part. The participating physics professors were briefed on the test's objectives after the 16 institutions agreed on treatment conditions for the treated and untreated groups. The professors were given a one-hour treatment session by the researchers on the use of ARSs and the preliminary conditions under which treatments must be implemented in order to improve proficiency and ensure that FA as well as polling technology are most likely used efficiently

(Fig. 3). A database of HOQs was also given to them for enhancing the ARSs tests (based on Higher-Order Questions). As a result of scheduling software, students were randomly assigned to one of two treatment groups, and three weeks later, classes were assigned by the researchers in a similar manner. A 31-week treatment cycle was followed by a 2-week testing period, a five-week summer break, and (the tests were excluded from the treatment). Based on the literature, more than twenty weeks are required for measuring the development of metacognitive skills (Butler, 2018). Before the test, participants completed the intrinsic motivation and metacognitive skills survey, and after the test, they completed a follow-up survey, one week later.

3.7. Procedure

This study's internal validity is increased by using a randomized experimental design with 16 participating colleges. For this, students from each college are randomly assigned by the scheduling application to one of the physics classes according to their educational level. Before the schedules are created, the data is randomly generated. The treatment is then randomly assigned within colleges based on draws to the treated and untreated groups, again based on educational level. We can assume that by randomizing classes, when dividing into the treatment and control groups, and becoming random did not rely on the professor, in this way, we reduce possible professor effects. We limit possible college effect and increase the external validity of our research, with including sixteen colleges in the treatment, by randomizing within these colleges.

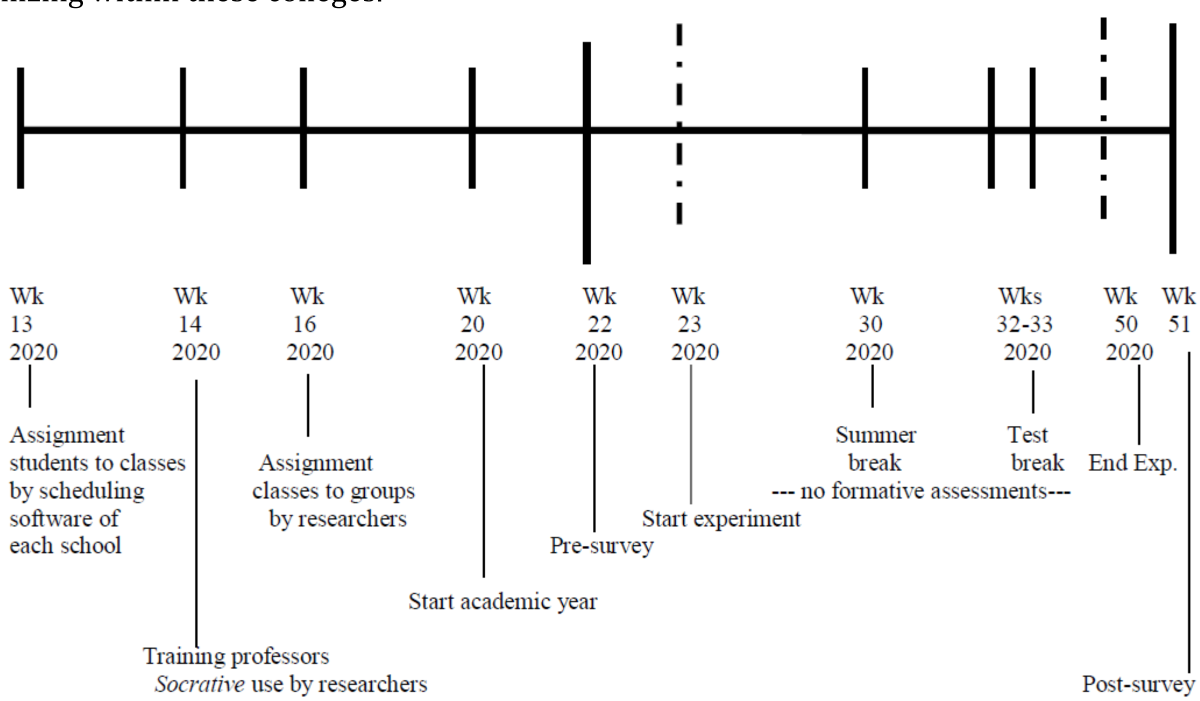

Fig. 3. Experimental timeline overview.

In addition, we reduce the risk of contamination, i.e. when students discuss HOQs outside of the classroom with one another. Students are required to take notes on college paper and leave them in the classroom at the end of each lesson, in order to eliminate this source of bias. In order to prevent any written HOQs material from being taken outside of the classroom, the professor's supervision is critical (Matthew & Ziemann, 2020).

3.8. Tools

Students are usually unaware of MS and IM because they are internal mental processes. Due to students' lack of awareness of cognitive system and intrinsic motivation (IM) processes, previous research shows that it is difficult for them to express their thoughts (Chng & Gurvitch, 2018; Chou, 2017; Howell et al., 2017; Mac George et al., 2008; Masania et al., 2018; Mayhew et al., 2020). Similarly, such mental conceptions are also not easy for researchers to assess (Mayhew et al., 2020; Paniagua & Swygert, 2016; Petto, 2019). Hence, we employ valid and trustworthy surveys to test these psychological constructs, the MS and IM.

3.8.1. Metacognitive skills

Students' MS are analyzed by the Heat and Temperature Metacognitive skills Awareness Inventory (HeTMAI) (Moh. Irma et al., 2021). This validated, self-reported questionnaire analyzes students' MS. HeTMAI consists of twenty-six items, classified into six psychological constructs: 1) Knowledge of Cognition (KoC); 2) Planning (PL); 3) Information Management (INFO); 4) Monitoring (MON); 5) Debugging (DEB); and 6) Evaluation (EVA) (See appendix C). On a 26 five-point Likert scale of 1 (Never) to 5 (Always), students provide their responses (Always). All five elements of the Likert scale are used to calculate the HeTMAI's total measure, which ranges from 1 to 5 (a higher ranking indicates a higher level of MS) (Moh. Irma et al., 2021; Pearson, 2019; Preszler, 2009; Russell, 2020). HeTMAI measures 'student's knowledge about oneself as a learner and about the learning' (KoC), and 'the knowledge of skills that allow students to plan their learning' (PL), 'how a student works on the information processes' (INFO), 'how a student applies skills of monitoring and regulation of its learning' (MON), 'what a student knows about learning and mastery of learning, specially, the learning achievement' (EVA).

3.8.2. Intrinsic Motivation

The Science Motivation Questionnaire II (SMQ-II) (Glynn et al., 2011) measures students' motivation to master physics by having them complete a motivation scale consisting of 25 five-point Likert scale items ranging from 1 (Never) to 5 (Always). The total measure for the SMQ-II is the average of all five-elements of Likert scale measures, with 1 being the lowest and 5 being the highest, a higher ranking denotes a greater degree of intrinsic motivation (IM) (Glynn et al., 2011; Pearson, 2019; Preszler, 2009; Russell, 2020).The both questionnaires i.e., HeTMAI and SMQ-II, which are translated into local languages by a panel of autonomous, expert and qualified professors, all experts in local languages and the native speakers in English. Next, the team of professors compared their work with the original English Models and worked on consensus building, on both questionnaires. After extensive discussions, professors resolved the inconsistencies and generalized joint Models in local languages. The questionnaires were translated back into English and compared to the original Models to see if they were consistent with the original English questionnaires. The surveys were eventually delivered to 48 students who were not research participants. They completed the questionnaires and commented on any items that were unclear.

3.9. Pre-survey records

All students were asked to complete a pre-survey that contained demographic information prior to the start of the treatment (ARSs use in the previous academic year, gender, physics test measures in the previous academic year, and age); as well as the pre-tests the Heat and Temperature Metacognitive skills Awareness Inventory (HeTMAI) (Moh. Irma et al., 2021) and Science Motivation Questionnaire (SMQ-II) (Glynn et al., 2011). The pre-survey was done in each college and class in week thirteen, one week before treatment began (Fig. 3). The professors explained the aim of the survey and urged students to take part, emphasizing that participation were entirely voluntary. All present students completed the survey voluntarily within a half-hour in the presence of the professor. Following that, all students who had been absent or sick during the class completed the pre-survey during the same week as the class. The amount of internal consistency for all contributing students as evaluated by Cronbach's alpha (Cronbach, 1951), is good for the pre-HeTMAI $\alpha = 0.86$. This alpha value exceeds the critical value of 0.70 (Cameron & Bizo, 2019). The same employs for the subscales; 1) Planning (PL; 05 statements: $\alpha = 0.81$); 2) Knowledge of Cognition (KoC; 06 statements: $\alpha = 0.88$); 3) Information Management (INFO; 04 statements: $\alpha = 0.85$); 4) Monitoring (MON; 04 items: $\alpha = 0.87$); 5) Debugging (DEB; 03 items: $\alpha = 0.90$); and 6) Evaluation (EVA; 04 statements: $\alpha = 0.85$). There are no significant differences between the Cronbach alpha coefficients and the previously reported HeTMAI values (Johns et al., 2012; Moh. Irma et al., 2021). Cronbach's alpha shows that the pre-SMQ-II coefficient of internal consistency for all individuals is an appropriate = 0.84, which is in line with Glynn's findings (2011). With the (aforementioned) critical value of 0.70 (Johns et al., 2012), we may assert that the pre-survey tools are valid and dependable.

3.10. Post-survey information

By completing the same survey and following the same protocols at the end of the treatment, researchers evaluated students' MS and IM. Just before winter break, we finished the post-survey (Fig. 3). The Cronbach's alpha measure (a= 0.89), is again good for answers to the post-HeTMAI. The same employs for the subscales; 1) Planning (PL; 05 statements: α = 0.81); 2) Knowledge of Cognition (KoC; 06 statements: α = 0.88); 3) Information Management (INFO; 04 statements: α = 0.85); 4) Monitoring (MON; 04 items: α = 0.87); 5) Debugging (DEB; 03 items: α = 0.90); and 6) Evaluation (EVA; 04 statements: α = 0.85). Cronbach's alpha shows that the pre-SMQ-II coefficient of internal consistency for all individuals is an appropriate = 0.87, which is consistent with Glynn's findings (2011). Based on the crucial value of 0.70 (Johns et al., 2012) from the (aforementioned) Cronbach alphas, we conclude that the post-survey tools are valid and reliable.

3.11. Data analysis

This treatment involved 966 students from 50 different classes, with 59% of them being female (41 percent boys), Table 1 summarizes the descriptive statistics. The students are an average of 17.90 years old (SD = 1.87), with ages ranging from 15 to 19, at the start of the treatment. The previous year's average physics test score was 19.35 (SD =.36), measured on a scale between 1 and 30. ARSs was previously used by 48% of the students, and the score is 3.48 (SD =.44). On a scale of 1 to 5, the average metacognitive skills (MS) measure is 4.55 (SD = 0.55). The average intrinsic motivation (IM) measure is 4.32 (SD = 0.64), again assessed on a scale between 1 and 5. Analysis of the MS and IM among the individual group, control group, as well as the boys and girls in the cooperative group, Table 2 presents. Using the analysis of variance (ANOVA), the effectiveness of randomization is evaluated. There was no significant difference between student groups (individual and cooperative) when analyzing data using ANOVA in comparison to control group. All of these student pre-treatment trends will be taken into account in our regressions in the next section. To aid in the interpretation of the findings, we'll use standardized (z-) measurements for all constructs.

We performed analysis of variance (ANOVA) in this study to examine if an individual or cooperative treatment affects MS and IM. Since ANOVA is the standard method for analyzing experimental data with two or more treatment conditions (Butler, 2018; Schell & Butler, 2018), following the treatment conditions, we adjust for students' observable pre-treatment components such as intrinsic motivation (IM), metacognitive skills (MS), age, gender, previous year ARS use, and physics exam measure. All standard errors in this analysis are clustered at the class level, since we randomized at the class level. To assess if the treatment has varied effects based on gender and metacognitive skill level, we include the interaction between the treatment condition and gender in Models 3 and 7, and the interaction between the treatment condition and metacognitive skills in Models 4 and 8. Finally, we apply a sequence of four regression analyses defined by Barchilon and Ben-Av (2016), for mediation effect testing. If (1) the treatment has a significant effect on the presumed mediator intrinsic motivation, (2) the treatment has a significant effect on the outcome construct metacognitive skills, and (3) the presumed mediator intrinsic motivation is significantly related to the outcome construct metacognitive skills, (4) we examine whether the effect of the professor regular descriptive feedback (RDF) or peer descriptive feedback (PDF) in combination with RDF on MS is reduced (partial-mediation) or no longer significant (complete mediation), if we control for the presumed mediator intrinsic motivation (IM).

4. RESULTS

4.1. Metacognitive skills

The effects of professor regular descriptive feedback (RDF; individual group) or peer descriptive feedback (PDF) with professor RDF (cooperative group) on MS are presented in Table 3. Regression analysis with clustered standard errors is used to estimate all models. Model one is a simple model with an outcome construct for MS, which solely refers to the student's treatment status. Model 1 shows that the cooperative treatment has a positive effect on MS when compared to the control group, with $\beta 2 = 0.54$ of SD, significance at $p < .001$. A combination of descriptive feedback from both peers and professors had a good and significant impact on the MS when compared to an untreated control group. The individual experiences a slight benefit, but nothing noteworthy. For each treatment condition, the mean post-measure of MS and 95% confidence intervals (CI) are shown in Figure 4. To demonstrate this, look at Figure 4, which shows that the CI between the control and cooperative groups do not overlap, but there is a significant overlap between the individual group and the other two groups. The cooperative group's mean is not only different from the control group's means, but it is also distinct from zero. This will increase the accuracy of our estimates and the explained variance in our model, and these constructs can also predict differences in outcomes (Temitayo, 2020). To find out how accurate our results are, we include pre-treatment constructs such as IM, MS, age, gender, use of Socrative (ARSs technology) during previous year, and physics test measure of the previous academic year in Model 2,. Incorporating these constructs marginally reduces the effect of CoT, but it remains significant at the $p < .001$ (with a standard deviation of $\beta 2 = 0.51$). Model 3 includes treatment indicator terms of interaction with gender to evaluate if treatment affects MS differently in boys and girls (Table 3). Considering that Model 3's reference group consists solely of males, its impacts on treatment status (InT or CoT) only reveal how effective the treatment is for boys, as opposed to how effective the treatment is for all genders.

Table 1. Descriptive Analysis.

	Sample	Mean	S D	Lower	Upper
Metacognitive skills (average score) pre-measure	966	4.55	.55	1	5
Intrinsic Motivation (average score) pre-measure	966	4.32	.64	1	5
Physics test measure previous year	966	19.35	.36	1	30
Socrative use previous year	966	2.48	.54	1	5
Age	966	17.90	1.87	15	19

Table 2. Pre-test scores of groups; Control group (CG), Individual group (InG), and Cooperative group (CoG), ANOVA results.

	CG (n = 222)		InG (n = 279)		CoG (n = 465)		F	P
	Mean	SD	Mean	SD	Mean	SD		
Metacognitive skills (average score) pre-measure	4.50	.53	4.45	.45	4.65	.65	1.79	.45
Intrinsic Motivation (average score) pre-measure	4.40	.43	4.05	.48	4.29	.59	.32	.92
Physics test score previous year	19.45	.33	19.25	.31	20.35	.46	.65	.68
Socrative use previous year	.89	.48	.84	.46	1.09	.68	.91	.52
Age	17.90	1.67	17.27	1.66	17.80	1.87	32.29	.76
Gender (girls)	.52	.55	.49	.48	.57	.52	1.65	.32

***p < 0.001, **p < 0.05, *p < 0.01.

Table 3. Standardized Post-measure metacognitive skills (MS), Regression analysis estimates.

	Model 1	Model 2	Model 3	Model 4
	MS Post-measure	MS Post-measure	MS post-measure	MS post-measure

(β_1) Individual Treatment (InT)[c]	.30 (.28)	.25 (.22)	- .030 (.19)	.093 (.21)
(β_2) Cooperative Treatment (CoT)[c]	.54*** (.18)	.51*** (.17)	.47*** (.15)	.37*** (.12)
(λ_1) Gender (Girl = 1)	-	-	-.37*** (.085)	-
(θ_1) Individual X Gender (0/1)	-	-	.67*** (.25)	-
(θ_2) Cooperative X Gender (0/1)	-	-	.24 (.21)	-
(λ_2) Metacognitive skills Low pre-measure (0/1)	-	-	-	- 2.030*** (.19)
(λ_3) Metacognitive skills Middle pre-measure (0/1)	-	-	-	- .47** (.19)
(η_1) InT X Metacognitive Low pre-measure (0/1)	-	-	-	.37 (.38)
(η_2) InT X Metacognitive Middle pre-measure (0/1)	-	-	-	.047 (.26)
(η_3) CoT X Metacognitive Low pre-measure (0/1)	-	-	-	.65** (.29)
(η_4) CoT X Metacognitive Middle pre-measure (0/1)	-	-	-	-.047 (.25)
(γ_k) Covariates Xi	No	Yes	Yes	Yes
Constant	-.31** (.15)	-.29** (.13)	-.38*** (.14)	.49** (.17)
Observations	966	966	966	966
R^2	.041	.37	.39	.35

***p < 0.001, **p < 0.05, *p < 0.01. Class-level standard errors are grouped together. [c] Compared with control group. Test measures of intrinsic motivation, metacognitive skills, and physics test scores, age, and use of Socrative are all control factors (tested at T0) for the experiment.

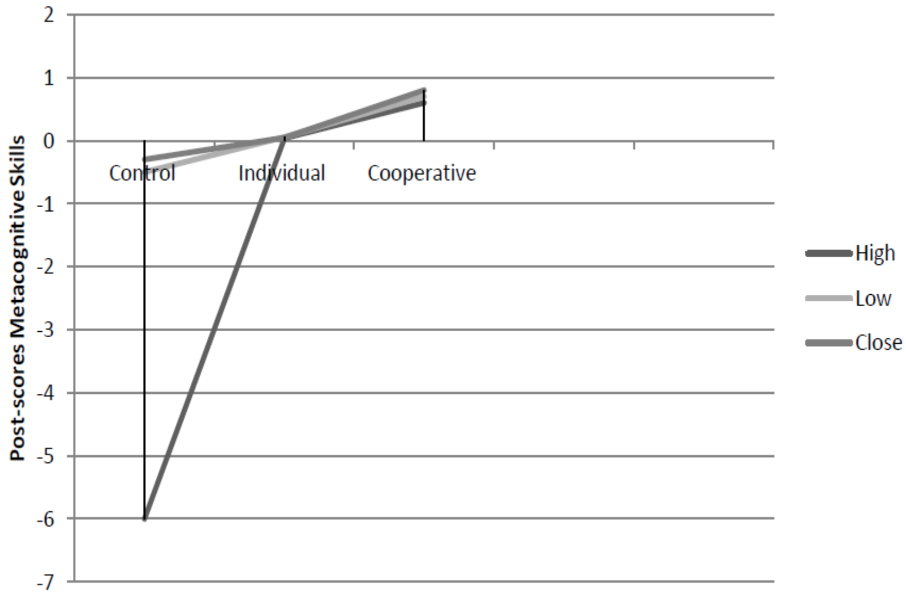

Figure 4. As a result of different treatment conditions, the mean standardized post-measures of metacognitive skills are depicted in Figure 4. In this case, zero represents the mean of all the students' standardized post-measure metacognitive skills.

Model 3 demonstrates a statistically significant discussion effect between gender and professor RDF (θ_1= 0.67, $p < .001$), meaning that girls who receive professor RDF outperform boys on post-measure MS. However, we do not detect any significant variations in a cooperative treatment between boys and girls (with a significant coefficient of .52 indicating that it is successful for both genders). Furthermore, girls in general measure score significantly lower on post-measure MS than boys (λ_1 = - .37, $p < .001$), whereas boys who receive cooperative treatment score significantly higher than boys who do not receive cooperative treatment (β_2 = .47, $p < .01$). We noted in the previous part that we divided all students into three MS subgroups (low, middle, and high) based on their MS pre-measure. Model 4 depicts the effects for students with high MS measures in comparison to the effects for the other students. Thus, the coefficients of the cooperative and individual treatments show the effects for students with high MS. In this scenario, the discussion impact is significant for students with a cooperative treatment and a low metacognitive pre-measure (η_3 = .65, $p < .05$), meaning that students with low MS benefit from this treatment much more than students with greater MS. Furthermore, as expected, students with low and moderate MS score significantly lower on their post-measure of MS than students with greater MS (λ_2 = - 2.03, $p < .001$; λ_3 = - .47, $p < .05$, respectively).

4.2. Motivation

Models 5 and 6 in Table 4 provide further information for the outcome IM. More specifically, Model 5 represents that the measure β_2 of CoT equals to .59 points of SD with significance level ($p < .001$). That is, PDF integrated with professor RDF has a significant and positive impact on IM. For the individual treatment, a non-significant impact for IM is determined. Fig. 5 shows the mean post-IM measurement together with 95% confidence intervals for each treatment condition. Despite the inclusion of covariance in the analyses to increase the precision of our estimates, the effects of the CoT in Model 6 remained with β_2 equal to.55 SD and a significance level (p.001). Model 7 incorporates treatment and gender considerations into the analyses in order to identify the heterogeneity of treatment results based on gender. In this situation, there is a significant positive interaction effect of treatment and gender (θ_1= 0.65, $p < .05$). As a result, even when using an individual group treatment, girls are more motivated than boys. The reference group is boys, meaning that girls in general do significantly lower on post measure IM than boys (λ_1 = - .36, $p < .01$), but boys who receive CoT perform significantly better than boys who do not receive CoT (β_2 = .45, $p < .05$). Model 8 shows the results of interaction analyses based on pre-measure MS of students for IM post-measure. It should be noted that students with high MS scores serve as a reference group for students with low and moderate MS scores. We opt to divide the groups according to pre-measure MS rather than pre-measure IM. This is because MS is our primary construct of interest, whereas motivation serves as a mediator and an intermediate outcome measure. We found no significant interaction effects for students with mild and moderate MS when compared to students with high MS in this study.

Table 4 Standardized Post-measure intrinsic motivation (IM), Regression analysis estimates.

	Model 5 IM Post-measure	Model 6 IM Post-measure	Model 7 IM post-measure	Model 8 IM post-measure
(β_1) Individual Treatment (InT)[c]	.090 (.29)	.088 (.27)	- .33 (.18)	.095 (.21)
(β_2) Cooperative Treatment (CoT)[c]	$.59^{***}$ (.17)	$.55^{***}$ (.15)	$.45^{**}$ (.14)	$.39^{***}$ (.12)
(λ_1) Gender (Girl = 1)	-	-	$-.36^{*}$ (.15)	-
(θ_1) Individual X Gender (0/1)	-	-	$.65^{**}$ (.24)	-
(θ_2) Cooperative X Gender (0/1)	-	-	.064 (.21)	-

(λ_2) Metacognitive skills Low pre-measure (0/1)	-	-	-	- 1.030 (.19)
(λ_3) Metacognitive skills Middle pre-measure (0/1)	-	-	-	.044 (.19)
(η_1) InT X Metacognitive Low pre-measure (0/1)	-	-	-	.037 (.22)
(η_2) InT X Metacognitive Middle pre-measure (0/1)	-	-	-	-.048 (.24)
(η_3) CoT X Metacognitive Low pre-measure (0/1)	-	-	-	.55 (.27)
(η_4) CoT X Metacognitive Middle pre-measure (0/1)	-	-	-	.049 (.24)
(γ_k) Covariates Xi	No	Yes	Yes	Yes
Constant	-.29 (.094)	.25 (.72)	-.28 (.18)	-.079 (.15)
Observations	966	966	966	966
R^2	.038	.57	.59	.57

***p < 0.001, **p < 0.05, *p < 0.01. Class-level standard errors are grouped together. [c] Compared with control group. Test measures of intrinsic motivation, metacognitive skills, and physics test scores, age, and use of Socrative are all control factors (tested at T0) for the experiment.

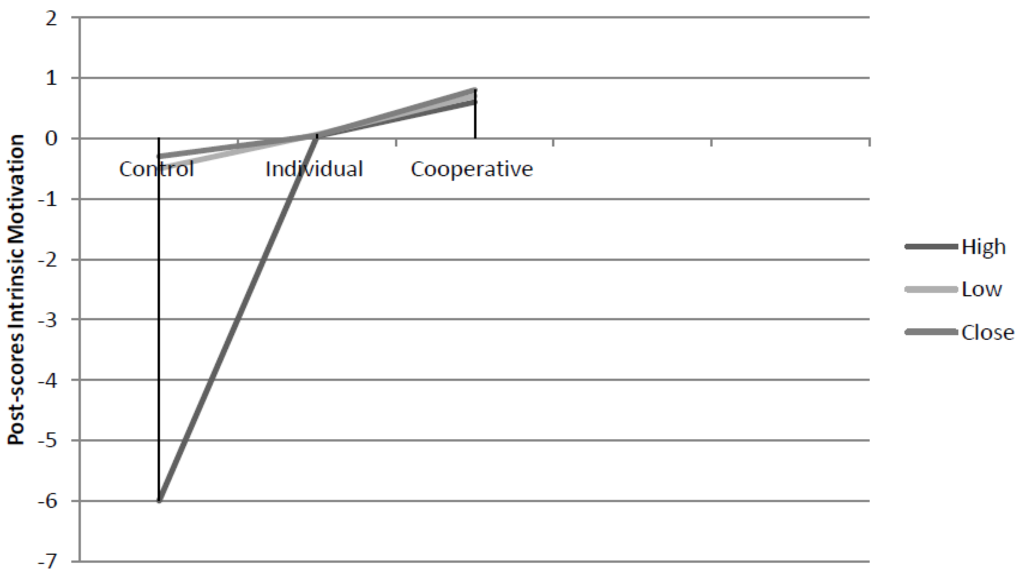

Figure 5. As a result of different treatment conditions, the mean standardized post-measures of intrinsic motivation are depicted in Figure 5. In this case, zero represents the mean of all the students' standardized post-measure metacognitive skills.

4.3. Mediation analyses

To assess whether IM mediates the projected effects of treatment status on MS, four regression analyses are undertaken. The first stage is to demonstrate a statistically significant relationship between treatment status (CoT or InT) and IM. Remember that in table 4, model 6 revealed that only the CoT had a statistically significant positive effect on IM ($\beta_2 = .55$, $p < .001$). As a result, we concentrate entirely on this considerable influence, which is represented in Fig. 6, track A. The second stage must be performed with MS treatment status. Remember that Model 2 in Table 3 shows that CoT has a significant positive influence on MS ($\beta_2 = .51$, $p < .001$). This effect is represented in Fig. 6 for Track B. The final step is to immediately include the IM post-measure into a regression with the outcome MS. Despite the fact that we do not adjust for treatment status, the findings of model 9 in table 5 demonstrate that IM is strongly associated with MS ($\gamma_6 = 0.89$, $p < 0.001$). This relationship is depicted in Fig. 6. As a result, IM can mediate the estimated effect of treatment status and MS, which can be observed if the IM post-measure is added to the treatment and MS post-measure regression analysis. Track D in Fig. 6, Model 10 in Table 5 show that the estimate β of CoT still strongly predicts MS ($\beta_2 = .84$, $p < .001$). According to the data, the impact size has been reduced slightly from 0.51 to 0.42, showing that IM partially mediates the effects of CoT on

MS. In addition, we use the Sobel test (Sobel, 1982). According to this test, when IM is incorporated into the models ($z = 3.65$, $p = .021$), the beta coefficients between a cooperative treatment and MS reduce significantly.

Table 5 Regression analysis predicts metacognitive skills for mediation analysis.

	Model 9 MS post-measure	Model 10 MS post-measure
(β_1) Individual Treatment (InT)[c]	-	.19 (.14)
(β_2) Cooperative Treatment (CoT)[c]	-	.42** (.098)
(γ_6) Motivation post-measure	.89*** (.074)	.84*** (.078)
(γ_k) Covariates Xi	Yes	Yes
Constant	-.73 (.74)	-.79 (.84)
Observations	966	966
R^2	.69	.71

***p < 0.001, **p < 0.05, *p < 0.01. Class-level standard errors are grouped together. [c] Compared with control group. Test measures of intrinsic motivation, metacognitive skills, and physics test scores, age, and use of Socrative are all control factors (tested at T0) for the experiment.

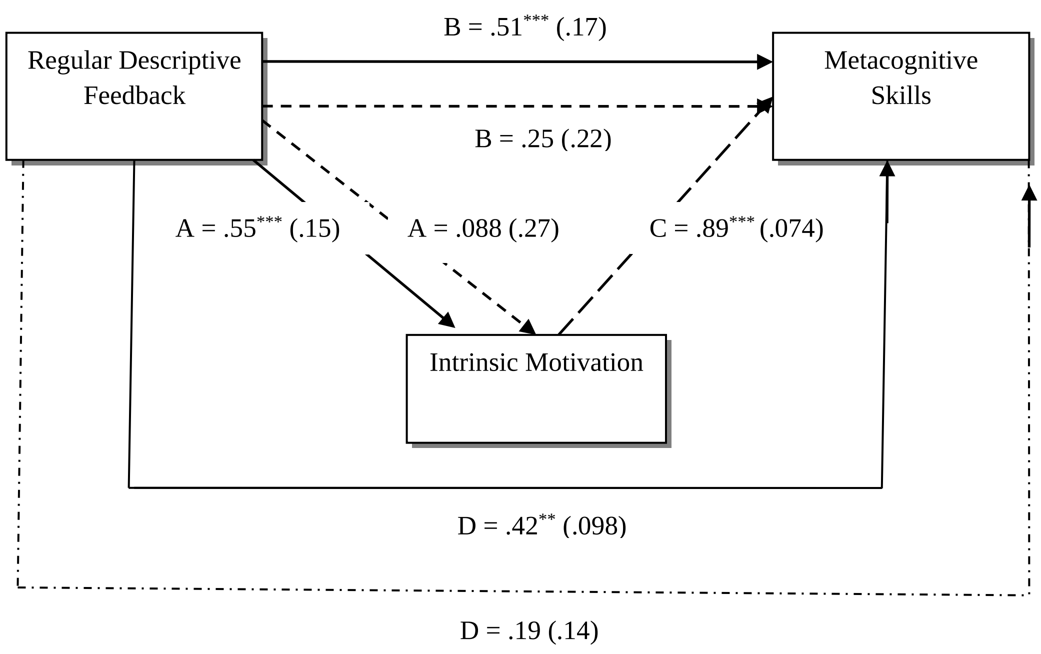

Fig. 6. Individual treatment (dashed arrows) and cooperative treatment (solid arrows), mediation analysis.

conforming that IM is a partial mediator here. Table C. 3 in appendix C, indicates the complete results of Table 5.

5. DISCUSSION

The purpose of this research is to investigate the impact of descriptive feedback (DF) during regular feedback techniques using ARSs technology on students' MS. As a result, we conducted a randomized field experiment (RFE) on 966 college students enrolled in physics programs in Pakistan. For intermediate outcomes intrinsic motivation (IM) and outcome MS, the consequences of descriptive feedback systems (e.g., RDF—professor regular descriptive feedback and/or PDF—peer descriptive feedback) are examined. As a result, our research study answers four research questions, are:

1. What impacts do professor regular descriptive feedback (RDF) or peer descriptive feedback (peer discussion) (PDF) integrated with professor RDF have on metacognitive skills (MS)?
2. What impacts do professor regular descriptive feedback (RDF) or peer descriptive feedback (peer discussion) (PDF) integrated with professor RDF have on intrinsic motivation (IM)?
3. Is there a difference in the effects for different student subgroups?
a. Do girls and boys have different outcomes?
b. Does it make a difference, if the student has a low, moderate, or high MS, if the effects differ?
4. Is the impact of professor RDF or PDF integrated with professor RDF, on students' MS mediated (in part) by its impacts on IM?

Regarding research questions 1 and 2, the findings indicate that PDF integrated with professor RDF have a significant positive impact on MS and IM. The results suggest that IM has a significant standardized impact size of about 0.6 and MS has a significant standardized effect size of about 0.6. The current findings are consistent with earlier studies (Brydges & Butler, 2012; Burton, 1991; Mazur, 1997) and corroborate the conclusions of Brady et al. (2013, 2015), Johns et al. (2012), and Mayer (2008), and Mayer et al. (2009) that ARSs technology should be merged with PDF to improve MS. Hunsu et al. (2016) reported a common outcome of approximately 0.8 SD for MS and 0.1 standard deviations SD for motivation in ARSs technology treatments, in their meta-analysis. It is difficult to compare effect sizes, however, because included studies only utilize control groups without ARSs technology or without HOQs, whereas in our study the control group also uses ARSs technology. According to our reporting, the majority of students (their discussion to professor) said, professor's explanations of proper reasoning are more efficient and constructive. We found this out through informal conversations with students. As a result, students demonstrate that they like group discussions, which break up the monotony of passive listening and guide them to support alternate solutions. Research by Suzan et al. (2010) and Masania et al. (2018) has also showed comparable outcomes. With PDF, students are more engaged and do not just take their instructors' or classmates' answers for granted. Since we used paired units of HOQs, we also recommend that PDF boost students' IM and MS. These talks let students learn from one another and put that learning to use when establishing follow-up HOQs. It fosters MS and reasoning, which makes it easier to transfer information and abilities from one problem to the next and from one context to another (Sajna & Premachandran, 2016; Zohar & Dori, 2012). Discussions of pairs of HOQs help students progress beyond the 'trial-and-error' problem-solving method.

We have seen that boys and girls are affected differently, in relation to the current research question 3a. Girls outperform boys on the post-measure MS and appear more motivated when professor RDF is employed as a motivational tool. Girls benefit from both individual (InT) and cooperative treatment (CoT) for MS and IM, but boys benefit only from a CoT for MS and IM. According to Compton and Allen (2018) and Deslauriers et al. (2019), both girls and boys benefit from ARSs systems' feedback techniques, but they approach learning in different ways. These gender impacts are in agreement. The results of this study are in line with earlier research, which suggests that girls may have a more positive attitude toward professor RDF and be more cognizant of the relevance of his feedback (Compton & Allen, 2018; Cubric & Jefferies, 2015; Flavell, 1976), which improves their perceptions of competence and MS (Deslauriers et al., 2019). We recommend combining peer discussion (PDF) and professor feedback (RDF) during a formative assessment with ARSs procedures to accommodate both girls and boys based on the findings of gender variances in MS and IM.

According to our findings in response to question 3b, there are also diverse effects across different MS subgroups of students. In comparison to students with high MS, those with low MS benefit much more by using PDF on top of professor RDF. Students with low MS may benefit more from a CoT than students with high MS, according to the findings of Vickery et al. (2015). Students with low MS may benefit more from contacts with high metacognitive classmates, which could explain our findings. Students with low metacognitive knowledge and skills benefit from instruction and collaboration with a more skilled metacognitive learner, according to David et al. (2015) and Suzan et al. (2020), this is consistent with our research study findings. In line with Vallely and Gibson's (2018) findings, we found that descriptive feedback (DF) reduces the cognitive load of students during learning, particularly for students with little MS who are confronted with more difficult problem-solving problems. Students with low MS have a greater need for HOQs with an 'analyzing' or 'evaluating' level because these HOQs require significantly less MS. This could explain some of our findings. We recommend teachers, based on these preliminary findings, to keep in mind the composition of the student group while interacting with peers; hence, students with strong metacognitive skills are partnered with those with low metacognitive skills.

Finally, in response to research question 4, we show that the impact of PDF combined with professor RDF on MS is mediated in part by the impact of IM. Because of this, formative assessment with peer discussion and professor RDF can motivate students, which in turn help to partially modify MS. A CoT and MS were found to be partially mediated by IM, which is a novel finding and contribution to the literature. In our opinion, PDF enables students to monitor their progress and improves their ability to assess how well they grasp academic material, both of which contribute directly to MS, this increases IM to study and as a result, increases MS. Additionally PDF makes lectures more lively and engaging, as well as making students more proactive and attentive when answering HOQs, which leads to increased IM, which indirectly increased MS. The more time spent during peer conversations on these HOQs, the better the results in MS may be. The important thing to remember, however, is that conceptualizing difficult cognitive processes take time and effort (Veenman & Elshout, 1995). The results of Vallely and Gibson (2018) and Mayhew et al. (2020) back up our claim that there would be no statistically significant increase in the MS, in this study.

Concerning the generalizability of the findings of this current study, we believe that these results can be applied in not only in the context of Pakistan but also will be effective in the other nations of the world, based on the two grounds. First, because our 16 colleges represent average colleges Pakistan, we can confidently assert that our findings apply to the vast majority of Pakistani institutions of secondary and upper secondary education. There is a 0.5 standard deviation difference between the mean of all constructs in the data for these colleges. Second, our HOQs cover all of the major academic topics of a normal college physics program (modern physics, thermodynamics, waves, electromagnetism, and mechanics); hence we believe the treatment is not context-bound. It also implies that there is no reason to believe that a comparable strategy would not be viable in different nations and cultures than Pakistan because these HOQs used in the treatments were generated by professors from the participating colleges (rather than by a context-specific academic author), they were completely based on the world-known revised Bloom's Taxonomy (Anderson & Krathwohl, 2001). When using audience response systems (ARSs; such as Socrative) in the cloud, we found that they have equivalent features and specifications (e.g., plotting their histograms) to other regularly used audience response systems, and so, the findings were almost certain to be identical. It is important to note that the findings can only be applied to countries where all students have digital devices and are used to keeping up with educational technology in class. The usage of data from multiple educational contexts is also required to implant knowledge that can be applied to the outcomes, as evidenced by the aforementioned findings.

6. CONCLUSION

Using the ARSs technology, this study shows that PDF followed by professor RDF improves students' MS, in part through IM. While students with low MS benefit more with PDF, students with high MS are more intrinsically motivated while using professor RDF, since girls have greater MS and are more intrinsically motivated in comparison to boys. Using randomized field experiment with a control group, this study confirms findings that formative assessments with ARSs technology should be used in conjunction with PDF and professor RDF to engage students in deeper metacognitive monitoring and development of MS (Brady et al., 2013; Brady et al., 2015; Brydges & Butler, 2012; Mazur, 1997).

Regardless of how well designed and executed this study was, there are still two issues that must be addressed in future studies. Even if a CoT improves MS symptoms, the true quality and amount of student-to-student interactions have yet to be studied. When it comes to student-to-student relationships, there are many variables that determine how much and how well students get along with each other. These variables include things like the time of day and the subject matter. Research in the future should concentrate on interactions and evaluate if the results gained were due to knowledge gains from interactions or to adopting responses from other students' responses. It's critical to conduct research in this area so educators can determine how much time they should give students to engage in effective discussions with their classmates, while yet keeping class time to a minimum. As a result, we were unable to answer questions about how feedback strategies increase student learning, even though we investigated the effects of descriptive feedback on students' MS while applying feedback techniques with ARSs technology. Future studies should look into how feedback tactics influence MS and whether instructors need to learn how to construct effective FA to help students understand the subject matter better. Many faculty use ARS technology for FA without considering the impact that feedback schemes can have on students' MS and their ability to study. Students will be examined utilizing formative assessments with ARSs technology while using PDF and professor RDF, which, based on present data, we can definitely recommend.

Conflict of interest

According to the authors, the research was carried out in the absence of any business or financial links that may be seen as capability warfare or hobby.

Acknowledgments

The authors would like to thank Dr. Khalid Jameel Rawat for providing helpful criticism on the manuscript's draft. We are appreciative of the talks with the three unnamed reviewers, Dr. Fawad Shams, Dr. Imtiaz Arif, and Dr. Naseem Qaisrani, as well as all participating students and professors from our participating colleges, who participated and provided their valuable feedback.

REFERENCES

Abdulla, M. H. (2018). The use of an online student response system to support the learning of Physiology during lectures to medical students. Education and Information Technologies, 23, 2931–2946. 10.1007/s10639-018-9752-0.

Abrahamson, L. (2006). A Brief History of Networked Classrooms: Impacts, Cases, Pedagogy, and Implications. In B. David (Ed.), Audience response systems in higher education: Applications and cases (pp. 1–25). Hershey, PA, USA: IGI Global.

Alexander, R. (2017). Towards Dialogic Teaching: Rethinking Classroom Talk.

Anderson, L. W., & Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. New York: Longman.

Armstrong ,N., Chang ,S. M., & Brickman M. (2007). Cooperative learning in industrial-sized biology classes.CBE Life Sciences Education, 6, 163–171. 10.1187/cbe.06-11-020017548878.

Astrid van T., Karel K., Marjan V., & Arnoud E. (2020). Professors' beliefs to integrate Web 2.0 technology in their pedagogy and their influence on attitude, perceived norms, and perceived behavior control. International Journal of Educational Research Open. https://doi.org/10.1016/j.ijedro.2020.1000=

Åsta H, Camilla B. & Magne D. (2018).Metacognitive skills in Language Learning and Teaching. Routledge 711 Third Avenue, New York, NY 10017.

Badia Valiente, J. D., Olmo Cazevieille, F., & Navarro Jover, J. M. (2016). On-line quizzes to evaluate comprehension and integration skills. Journal of Technology and Science Education, 6, 75–90. 10.3926/jotse.189.

Baker, Philip R. A., Francis, Daniel P., Cathcart, & Abby (2017). A Mock Randomized Controlled Treatment With Audience Response Technology for Teaching and Learning Epidemiology. Asia-Pacific Journal of Public Health. 29 (3): 229–240. doi:10.1177/1010539517700473. PMID 28434251.

Bale, D. (2018) 'This Norfolk church is using a phone app to rate hymns', North Norfolk News, 4 Dec. [online] Available at: https://www.northnorfolknews.co.uk/news/aylsham-church-treatmentling-use-of-a-live-polling-smartphone-app-1-5805737.

Barchilon Ben-Av, M., & Ben-Av, R. (2016). Digital devices: From distraction to attraction. Journal of Educational Technology Systems, 45, 93–102. 10.1177/0047239516639720.
Barber M, & Njus D. (2007). Clicker evolution: seeking intelligent design. CBE Life Science Education; 6(1):1-8. 8.
Beatty, I. D., Grace, W. J., Leonard, W. J. & Dufense R. J. (2006). Designing effective questions for classroom response systems teaching. American Journal of Physics, 74(1), 31–39. 10.1119/1.2121753.
Bowen, J.A. (2012). Teaching naked: how moving technology out of your, the classroom will improve student learning. San Francisco, CA: Jossey-Bass
Brady M., Seli H., & Rosenthal J. (2013). “Clickers” and metacognitive skills: A quasi-experimental comparative study about metacognitive self-regulation and use of electronic feedback technology. Computers & Education, 65, 56–63. 10.1016/j.compedu.2013.02.001
Brady M, Rosenthal J, & Forest CP (2015). The effect of audience response systems on metacognitive skills in graduate students: a two-year mixed method study. Presented at: Innovations in Medical Education Conference; February 21, 2015; San Gabriel, CA
Bruff, D. (2009). Teaching with Classroom Response Systems: Creating Active Learning Environments. San Francisco : Jossey-Bass, ©2009.
Brydges, R., & Butler, D. (2012). A reflective analysis of medical education research on self-regulation in learning and practice. Med Educ; 46 (1):71-79.
Butler, A. C. (2018). Multiple-choice testing in education: Are the best practices for assessment also good for learning? Journal of Applied Research in Memory and Cognition, 7(3), 323-331.
Burton, S. J. (1991). How to Prepare Higher Multiple-Choice Test Items: Guidelines for The Faculty. Brigham Young The Testing Services and The Department of Instructional Science.

Cain J. Robinson E. (2008). A primer on audience response systems: Current applications and future considerations. American Journal of Pharmaceutical Education, 72(4), 77. 10.5688/aj72047719002277

Cameron, K. & Bizo, L. (2019) 'Use of the game-based learning platform KAHOOT! to facilitate learner engagement in animal science students', Research in Learning Technology, vol. 27 [online] Available at: https://doi.org/10.25304/rlt.v27.2225

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. 10.1002/j.2168-9830.2010.tb01052.

Chi M. T. H. De Leeuw N. Chiu M. & LaVancher C. (1994). Eliciting self explanations improves metacognitive skills. Cognitive Science, 18, 439–477.

Chng, L.; & Gurvitch, R. (2018). Using Plickers as an Assessment Tool in Health and Physical Education Settings. Journal of Physical Education, Recreation and Dance , 89 (2), 19– 25, DOI: 10.1080/07303084.2017.=04510

Chou, C.-C. (2017). An Analysis of the 3D Video and Interactive Response Approach Impacts on the Science Remedial Teaching for Fourth Grade Underachieving Students. EURASIA Journal of Mathematics, Science and Technology Education, 13 (4), 1059– 1073, DOI: 10.12973/eurasia.2017.00658a

Coleman E. B. (1998). Using explanatory knowledge during collaborative problem solving in science. Journal of the Learning Sciences, 7, 387–427. 10.1080/10508406.1998.9672059

Collins J. (2008). Audience response systems: Technology to engage learners. Journal of the American The of Radiology, 5, 993–1000. 10.1016/j.jacr.2008.04.00818755440

Collins L. J. (2007). Livening up the classroom: using audience response systems to promote active learning. Medical Reference Services Quarterly, 26, 81–88. 10.1300/J115v26n01_0817210552.

Compton, M. & Allen, J. (2018) 'Student response systems: a rationale for their use and a comparison of some cloud-based tools', Compass: Journal of Teaching and Learning, vol. 11, no. 1 [online] Available at: https://pdfs.semanticscholar.org/5f77/f5e529996962df3c15bf6ff1e4d97f48fb88.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.

Crossgrove K. & Curran K. L. (2008). Using clickers in non-majors- and majors-level biology classes: student opinion, learning, and long-term retention of course material. CBE Life Sciences Education, 7, 146–154. 10.1187/cbe.07-08-006018316817

Crouch, C. H., & Mazur, E. (2001). Peer Instruction: Ten years of experience and results. American Journal of Physics, 69(9), 970–977. https://doi.org/10.1119/1.1374249.

Crowe, A., Dirks, C., & Wenderoth, M. P. (2008). Biology in bloom: Implementing Bloom's taxonomy to enhance student learning in biology. CBE-Life Sciences Education, 7(4), 368–381.

Cotton, S. M., Kiely, P. M., Crewther, D. P., Thomson, B., Laycock, R., & Crewther, S. G. (2005). A normative and reliability study for the Raven's Coloured Progressive Matrices for primary school aged children from Victoria, Australia. Personality and Individual Differences, 39, 647– 659. https://doi.org/10.1016/j.paid.2005.02.015.

Cubric, M. & Jefferies, A. (2015) The benefits and challenges of large-scale deployment of electronic polling systems: The student views from across varied subject groups, Computers & Education, 87, pp.98-11'1. doi:10.1016/j.compedu.2015.04.004.

Deslauriers, L. et al. (2019) 'Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom', Proceedings of the National Academy of Sciences of the United States of America, vol. 116, 19251–19257 [online] Available at: https://www. pnas.org/content/116/39/19251

David, W. C., Matthew F., & Carolyn L., (2015). Read The Room For Real: How a Simple Technology Creates Higher Meetings. I AM Publications; 2nd edition

De Corte E, Verschaffel L, & Masui C (2004). The CLIA-model: a framework for designing powerful learning environments for thinking and problem solving. European Journal of Psychology Education; 19(4):365-384. 7.

Dembo M. & Seli H. (2007). Motivation and learning strategies for the success: A self-control approach (3rd ed.). New York, NY: Taylor & Francis.

Draper S. W. & Brown M. K. (2004). Increasing interactivity in lectures using an electronic polling system. Journal of Computer Assisted Learning, 20, 81–94. 10.1111/j.1365-2729.2004.00074.x

Eva KW, & Regehr G (2011). Exploring the divergence between self-assessment and self-monitoring. Advances in Health Science Education; 16(3): 311-329. 10.

Flavell J. H. (1976). Metacognitive aspects of problem solving. In ResnickL. B. (Ed.), The nature of intelligence (pp. 231–236). Hillsdale, NJ: Erlbaum.

Flavell J. H. (1979). Metacognitive skills and cognitive monitoring: A new area of cognitive-developmental inquiry. The American Psychologist, 34, 906–911. 10.1037/0003-066X.34.10.906

Flavell J. H. (1987). Speculation about the nature and development of metacognitive skills. Metacognitive skills, motivation, and metacognitive skills (pp. 21–29). Hillsdale, NJ: Lawrence Erlbaum.

Forest CP (2012). The effect of audience response systems on adult learning: evidence-based rationale and ARS implementation guide. Journal of Physician Assist Education; 23(4):54-59. 2.

Glynn, S., M., Brickman, P., Armstrong, N., & Gita, T., (2011). Science motivation questionnaire II: Validation with science majors and non-science majors. Journal of Research in Science Teaching. https://doi.org/10.1002/tea.20442

Grahame JA (2016). Digital note-taking: discussion of evidence and best practices. Journal of Physician Assist Education; 27(1):47-50. 3.

Guarascio, A., Nemeck, B. & Zimmerman, D. (2017) 'Evaluation of students' perceptions of the ARS application verses a traditional student response system and its impact on classroom engagement', Currents in Pharmacy Teaching and Learning, vol. 9, no. 5, pp. 808–812. doi: 10.1016/j.cptl.2017.05.011

Hake, R. (1998) 'Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics classes', American Journal of Physics, vol. 66, no. 1, pp. 64–74. doi: 10.1119/1.18809

Herreid C. F. (2010). It's all their fault. Journal of Microbiology and Biology Education, 11, 34–36. 10.1128/jmbe.v11i1.13823653695

Heaslip, G., Donovan, P. & Cullen, J.G. (2014) 'Student response systems and learner engagement in large classes', Active Learning in Higher Education, vol. 15, no. 1, pp. 11–24. doi: 10.1177/=697874135=648

Hung, H. (2016) 'Clickers in the flipped classroom: bring your own technology (BYOD) to promote student learning', Interactive Learning Environments, vol. 25, no. 8, pp. 983–995. doi: 10.1080/10494820.2016.1240090

Howell, D. D.; Tseng, D. C.; & Colorado-Resa, J. T. (2017). Fast Assessments with Digital Tools Using Multiple-Choice Questions. The Teaching , 65 (3), 145– 147, DOI: 10.1080/87567555.2017.129=89

Hunsu NJ, Adesope O, & Bayly DJ (2016). A meta-analysis of the impacts of audience response systems (clicker-based Technology) on cognition and affect. Computers & Education; 94:102-119. 4.

Johns, B.T., Thomas, M. G., David, B. P., & Michael, N. J., (2012). Effects of word frequency, contextual diversity, and semantic distinctiveness on spoken word status. The Acoustical Society of America 132; doi: 10.1121/1.4731641. https://doi.org/10.1121/1.4731641

Karpicke, J. D. (2012). Retrieval-Based Learning: Active Retrieval Promotes Meaningful Learning. Current Directions in Psychological Science, 21(3), 157–163.

Knight, J. & Wood, W. (2005) ‘Teaching more by lecturing less’, Cell Biology Education, vol. 4, no. 4, pp. 298–310 [online] Available at: http://doi.org/10.1187/05-06-0082 E.

Krause, J. M.; O’Neil, K.; & Dauenhauer, B. (2017). Plickers: A Descriptive Assessment Tool for K–12 and PETE Professionals. Strategies, 30 (3), 30– 36, DOI: 10.1080/08924562.2017.1297751

Little, C. (2016). ‘Technological review: Mentimeter smart phone student response systems’, Compass: Journal of Learning and Teaching, vol. 9, no. 13 [online] Available at: https:// journals.gre.ac.uk/index.php/compass/article/view/328/pdf

MacGeorge E. L. Homan S. R. Dunning J. B. Elmore D. Bodie G. D. & Evans E. (2008). Student evaluation of audience response technology. Educational Technology Research and Development, 56, 125–145. 10.1007/s1=23-007-9053-6

Masania, J.; Grootveld, M.; & Wilson, P. B. (2018). Teaching Analytical Chemistry to Pharmacy Students: A Integrated, Iterative Approach. Journal of Chemical Education , 95 (1), 47– 54, DOI: 10.1021/acs.jchemed.7b00495

Matthew McKenzie, & M. Ziemann (2020). Assessment of the web-based audience response system ARS for biomedical science revision classes. International Journal of Educational Research Open. https://doi.org/10.1016/j.ijedro.2020.100008 .

Mayhew, E., Davies, M., Millmore, A., Thompson, L. & Pena, A. (2020) The impact of audience response platform Mentimeter on the student and staff learning experience. Research In Learning Technology, 28. ISSN 2156-7077 doi: https://doi.org/10.25304/rlt.v28.2397 Available at http://centaur.reading.ac.uk/91682/

Mayer R. E. (2008). Learning and instruction. Pearson Education, Inc.

Mayer R. E. Stull A. DeLeeuw K. Almeroth K. Bimber B. & Chun D. (2009). Clickers in the classrooms: Fostering learning with questioning methods in large lecture classes. Contemporary Educational Psychology, 34, 51–57. 10.1016/j.cedpsych.2008.04.002

Mazur E. (1997). Peer instruction: A user’s manual. Upper Saddle River, NJ: Prentice Hall.

Melanie B & Christopher P Forest (2018). Metacognitive skills, Descriptive Assessment, and Student Perspective: Learning About Metacognitive skills Through In-Class Comparison of Response Systems. Journal of Physician Assistant Education. DOI: 10.1097/JPA.0000000000000203.

Moh. Irma, Dwi Sulisworo, Jumadi, Heru Kuswanto, & Siti Anisatur R. (2021). Heat and temperature metacognitive skills awareness inventory: A confirmatory factor analysis. International Journal of Evaluation and Research in Education. 10:2, pp. 389-395

Mountford-Zimdars, A., Sanders, J., Moore, J., Sabri, D., Jones, S. & Higham, L. (2017) What can universities do to support all their students to progress successfully throughout their time at the? Perspectives: Policy and Practice in Higher Education, 21(2-3), pp.101-110.

Mshayisa, V. V. (2020). Students' Perceptions of Plickers and Crossword Puzzles in Undergraduate Studies. Journal of Food Science and Education , 19, 49– 58, DOI: 10.1111/1541-4329.12179

Paniagua, M. A., & Swygert, K. A. (Ed.) (2016). Constructing Written Test Questions For the Basic and Clinical Sciences (4th ed.). National Board of Medical Examiners (NBME), Philadelphia, PA

Pearson, R. J. (2019). Exploring Peer Instruction: Should Cohort Clicker Responses Appear During or After ARSs? Journal of Chemical Education, 96 (5), 873– 879, DOI: 10.1021/acs.jchemed.9b00035

Petto, A. J. (2019). Technology Meets Pedagogy: Comparing Classroom Response Systems. Journal of The Science Teaching. 48(4), (Mar/April 2019): 55-63.

Perera, V. H., & Hervás-Gómez, C. (2021). The students' perceptions toward the use of an online student response system in game-based learning experiences with mobile technology. European Journal of Educational Research, 10(2), 1009-1022. https://doi.org/10.12973/eu-jer.10.2.1009

Pink, D. H. (2009). Motivate: The surprising truth about what motivates us. New York

Preszler R. W. (2009). Replacing lecture with peer-led workshops improves student learning.CBE Life Sciences Education, 8, 182–192. 10.1187/cbe.09-01-000219723813

Reay, N. W., Li P., & Bao, L. (2008). Testing a new polling system question methodology. American Journal of Physics, 76, 171–178. 10.1119/1.2820392

Ruth, W. & Shaista, S. (2020). A systematic review of audience response systems for teaching and learning in higher education: The student experience. The Computers & Education. 153:4

Russell J. (2020). Clickers versus Plickers: Comparing Two Audience Response Systems in a Smartphone-Free Teaching Environment. J. Chem. Educ. 2020, 97, 8, 2342–2346

https://doi.org/10.1021/acs.jchemed.0c00464
Sajna J., & Premachandran. P (2016). A Study on the Metacognitive Awareness of Secondary College Students. Universal Journal of Educational Research 4(1): 165-172. DOI: 10.13189/ujer.2016.040121
Schell, J. A., & Butler, A. C. (2018). Insights From the Science of Learning Can Inform Evidence-Based Implementation of Peer Instruction. Frontiers in Education. https://doi.org/10.3389/feduc.2018.00033
Suzan O., James De., & Eglė B., (2020). Classroom Response Systems: Inclusion, Attainment and Active Learning. New Directions in the Teaching of Physical Sciences. 15:3.
Temitayo, D. O. (2020). Sudden change of pedagogy in education motivationn by COVID-19: Perspectives and evaluation from a developing country. Research in Globalization. http://creativecommons.org/licenses/by-nc-nd/4.0/.
Vallely, K. & Gibson, P. (2018) 'Engaging students on their technology with Mentimeter', Compass Journal of Learning and Teaching, vol. 11, no. 2 [online] Available at: https://journals.gre. ac.uk/index.php/compass/article/view/843/pdf
Veenman M. V. J. & Elshout J. J. (1995). Variedial impacts of instructional support on learning in simulation environments. Instructional Science, 22, 363–383. 10.1007/BF00891961.
Vickrey, T., Rosploch, K., Rahmanian, R., Pilarz, M., & Stains, M. (2015). Based implementation of peer instruction: A literature review. CBE—Life Sciences Education, 14(1), es3.
Ward, A. F.; Duke, K.; Gneezy, A.; & Bos, M. W. (2017). Brain Drain: The Mere Presence of One's Own Smartphone Reduces Available Cognitive Capacity. J. Assoc. Consum. Res. , 2 (2), =0– 154, DOI: 10.1086/69=62
Wengrowicz, N., Dori, Y. J., & Dori, D. (2018). Metacognitive skills and meta-assessment in engineering education. In Y. J. Dori, Z. Mevareach, & D. Bake (Eds.), Cognition, metacognitive skills and culture in STEM education. Springer.
Williams, N., Hussain, H., Manojkumar, P. & Thapa, A. (2016) An evaluation of a STEM summer undergraduate research internship scheme: student-perceived learning gains. New Directions in the Teaching of Physical Sciences, vol. 11. doi:10.29311/ndtos.v0i11.581
Yee, S.L.W. & Ean, C.L.C. (2020) 'Malaysian private the students' perception of online discussion forums: a qualitative enquiry', Sains Humanika, vol. 12, no. 2 [online] Available at: https://sainshumanika.utm.my/index.php/sainshumanika/article/view/1610

Zimmerman B. J. (2000). Attaining self-regulation: A social cognitive perspective. In Boekaertz M.PintrichP. R. Zeidner M. (Eds.), Handbook of self-regulation (pp. 13–19). San Diego, CA: Academic Press. 10.1016/B978-012109890-2/50031-7

Zohar, A., & Dori, Y. J. (Eds.). (2012). Metacognitive skills in science education: Trends in current research. Dordrecht: Springer.