Chemistry Education Research and Practice

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Chem. Educ. Res. Pract.*, 2024, **25**, 833

Augmented reality meets Peer instruction†

Daniel Elford, *[a] Garth A. Jones[b] and Simon J. Lancaster *[c]

Peer Instruction (PI), a student-centred teaching method, engages students during class through structured, frequent questioning, facilitated by classroom response systems. The central feature of PI is the ConcepTest, a question designed to help resolve student misconceptions around the subject content. Within our coordination chemistry PI session, we provide students two opportunities to answer each question – once after a round of individual reflection, and then again after a round of augmented reality (AR)-supported peer discussion. The second round provides students with the opportunity to "switch" their original response to a different answer. The percentage of right answers typically increase after peer discussion: most students who answer incorrectly in the individual round switch to the correct answer after the peer discussion. For the six questions posed, we analysed students' discussions, in addition to their interactions with our AR tool. Furthermore, we analyse students' self-efficacy, and how this, in addition to factors such as ConcepTest difficulty influence response switching. For this study, we found that students are more likely to switch their responses for more difficult questions, as measured using the approach of Item Response Theory. Students with high pre-session self-efficacy switched from right-to-wrong ($p < 0.05$) and wrong-to-different wrong less often, and switched from wrong-to-right more often than students with low self-efficacy. Students with a low assessment of their problem solving and science communication abilities were significantly more likely to switch their responses from right to wrong than students with a high assessment of those abilities. Analysis of dialogues revealed evidence of the activation of knowledge elements and control structures.

Received 1st May 2023,
Accepted 13th April 2024

DOI: 10.1039/d3rp00093a

rsc.li/cerp

Open Access Article. Published on 19 April 2024. Downloaded on 5/10/2026 12:06:01 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Introduction

Visualising chemistry concepts can be a challenging, but crucial skill for the fledgling chemist. To alleviate this challenge, we present an educational intervention that affords impactful augmented reality (AR) experiences synergised with the pedagogical approach of Peer Instruction (PI). AR is a technique that superimposes computer-assisted contextual information onto the physical world (Milgram *et al.*, 1995), obviating the reliance upon using 2D representations of 3D phenomena. For this study, participants used an AR tool called ChemFord (Elford *et al.*, 2022, 2023). ChemFord is a free AR mobile and tablet application available on Apple iOS (11.0 or later) and Android (4.4 and up) platforms. This educational intervention, utilising the student-centred approach of PI alongside AR-supported discussions for teaching topics of coordination chemistry, is the first published (to our knowledge). The underlying hypothesis of this work is that thinking about the qualitative dialogue, in terms of activated resources supported by AR, alongside the quantitative measures, in terms of ConcepTest difficulty and self-efficacy, can give insights into how conceptual development takes place.

Within a PI session, time is organised by a sequence of questioning, interactive discussion, and explanation (Schell and Mazur, 2015). The element of peer discussion is arguably the most recognizable feature of the PI model, and works to maximise both the amount of time that students think about key concepts, in addition to the time students spend engaging in self-monitoring of their understanding of the discipline. As students explain their understanding of a ConcepTest, often an epiphany occurs, which takes them further than their individual thinking processes. The body of research on PI, primarily from physics education researchers indicates that PI significantly improves student learning outcomes, such as conceptual understanding and problem-solving ability. As such, implementation of the process outlined in Fig. 1 has provided compelling evidence that PI is associated with substantial improvements in students' ability to solve conceptual and quantitative problems (Mazur, 1997; Vickrey *et al.*, 2015).

[a] School of Chemistry, University of East Anglia, Norwich Research Park, NR4 7TJ, UK. E-mail: d.elford@uea.ac.uk
[b] School of Chemistry, University of East Anglia, Norwich Research Park, NR4 7TJ, UK. E-mail: garth.jones@uea.ac.uk
[c] School of Chemistry, University of East Anglia, Norwich Research Park, NR4 7TJ, UK. E-mail: s.lancaster@uea.ac.uk

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3rp00093a

Open Access Article. Published on 19 April 2024. Downloaded on 5/10/2026 12:06:01 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

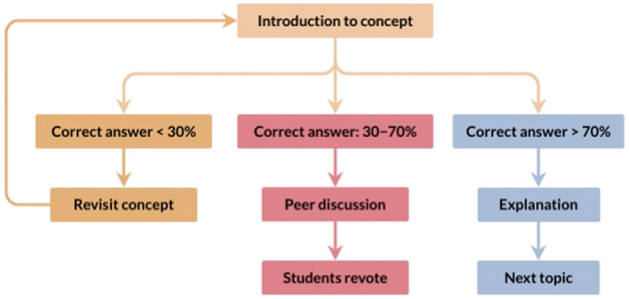


Fig. 1 PI implementation procedure, adapted from Mazur (1997).

Theoretical approach

Within this study, we focus on the affective domain factor of perceived self-efficacy, an individual's belief that one can successfully complete a task (Bandura, 1977). Self-efficacy has been previously recognised as a strong predictor of performance in science (Andrew, 1998; Pietsch *et al.*, 2003; Dogan, 2015). Students with higher self-efficacy are reported to experience fewer negative emotions in the face of difficulty, compared to students with lower reported self-efficacy (Bartimote-Aufflick *et al.*, 2015). Self-efficacy has been shown to influence cognition, motivation, and affective processes, which in turn, can influence future self-efficacy beliefs (this is known as reciprocal determinism, Bandura, 1977). People are not simply acted upon by external forces, but rather choose to place themselves in environments that they believe are conducive for their learning. Past experiences will influence reinforcements, expectations, and expectancies, all of which shape whether, and why, a person engages in specific behaviours.

Self-efficacy was first developed as an integral part of social cognitive theory (SCT), an agentic perspective to human development, adaptation, and change. As there are different social cognitive theoretical perspectives, the focus of this study is limited to the social cognitive theory proposed by Bandura (1986, 1997, 2001). SCT posits that learning occurs in a social context with a dynamic and reciprocal interaction of the person, environment, and behaviour (Bandura, 1986). Within this triadic reciprocality, each set of influences on human functioning affects the others, and is in turn affected by them. The pivotal feature of SCT is the importance of social influence, and its emphasis on external and internal social reinforcement.

The construct of self-efficacy within SCT refers to the level of a person's confidence in his or her ability to successfully perform an action. Thus, to support the establishment of perseverance and self-regulated learning within our PI environment, students were randomly organised into groups of 2–3 individuals. This allowed students to support one another, whilst making their thinking explicit through discussion. Social cognitive theorists emphasize that learning is most effective when peers learn from others, who are both similar to themselves, and display high levels of self-efficacy (Schunk, 2005). For example, students who feel competent about performing well in mathematics (high self-efficacy) are apt to engage in effective learning strategies that will benefit their learning (behavioural), as well as demonstrating greater persistence (Schunk and DiBenedetto, 2016; Schunk and Usher, 2019).

Meta-Analyses have been conducted on studies with diverse experimental and analytical methodologies applied across diverse spheres of functioning (Boyer *et al.*, 2000; Moritz *et al.*, 2000; Stajkovic *et al.*, 2009). The accumulated evidence confirms that efficacy beliefs contribute significantly to the quality of human functioning. Cognitively, our intention was that the 3D perspective afforded by ChemFord would help manage working memory load, in addition to providing insight into the structure.

ConcepTest development

The centrepiece of PI is the ConcepTest; a question designed to assess students' understanding of the principal concepts underlying the learning material (Mazur, 1997; Lancaster *et al.*, 2019). Within a low-stakes environment, ConcepTests promote higher-order thinking, allowing students to demonstrate cognitive skills that are conduits to learning. ConcepTests also give students extensive retrieval practice (Halpern and Hakel, 2003), the act of generating the same information in different applications to promote long-term memory. As such, ConcepTests may be considered equivalent to comprehension, application, or analysis questions as defined by Bloom's Taxonomy (McConnell *et al.*, 2003).

Smith *et al.* (2009) report that students improve the most when asked difficult questions during PI, a trend that was also found by Porter *et al.* (2011). In addition, lower learning gains have also been reported for instructors implementing easier ConcepTests (Rao and DiCarlo, 2000; Knight *et al.*, 2013). Hence, empirical evidence suggests that the benefits of PI, especially the effectiveness of student discussions, is very likely influenced by the difficulty of the question posed. In their longitudinal analysis, Crouch and Mazur (2001) found that substantial learning gains following voting in round 2 (post-discussion) occurred when the voting in round 1 was correct for 35–70% of the student base. Below 35%, the concept may still be too alien, requiring the provision of further description (Simon *et al.*, 2010).

As such, we developed six ConcepTests to probe students' comprehension of organometallic chemistry concepts (see ESI,† for details of ConcepTests). Throughout the development process, internal validation with experts in the field of inorganic chemistry at UEA was carried out to ensure student attention was focused towards critical concepts key to addressing specific learning goals. To satisfy these requirements, we used the following six criteria for each ConcepTest (Newbury, 2013):

i. **Clarity.** Students should waste no cognitive resources understanding the requirements of the question.

ii. **Context.** The question should be appropriate for the learning material.

iii. **Learning outcome.** The question should allow students to demonstrate that they grasp the concept.

iv. **Distractors.** Distractors should be plausible solutions to the question.

This journal is © The Royal Society of Chemistry 2024

Open Access Article. Published on 19 April 2024. Downloaded on 5/10/2026 12:06:01 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

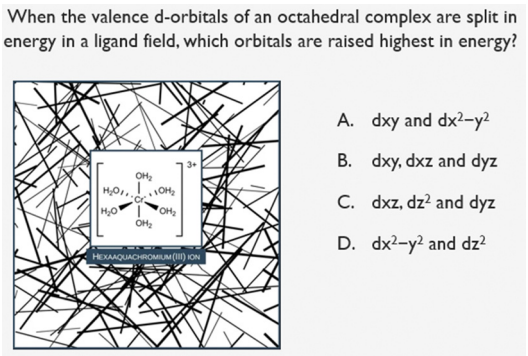


Fig. 2 The first (of six) ConcepTests developed for our PI session.

v. **Difficulty.** The question should not be too easy or too hard.

vi. **Stimulates thoughtful discussion.** The question should engage students, and incentivise thoughtful discussion.

Regarding the implementation of AR technology into PI, a very limited number of previous works are reported (Ravna *et al.*, 2022; Themelis, 2022). Although VR is commonly preferred for multiuser collaboration, the role of AR for collaboration is increasing. As such, throughout ConcepTest development, we focused on how the affordances of AR could be leveraged to promote important discussion points. In Fig. 2, we present our first ConcepTest. To answer the first ConcepTest correctly, there are three conceptual points which, fundamentally, students must understand:

i. Firstly, students must recognise how the axial and equatorial aqua ligands are situated around the chromium metal atom.

ii. Secondly, students must be able to comprehend the shapes and orientations of the five d-orbitals of the chromium metal atom.

iii. Lastly, students must be able to comprehend the consequence of ligand and chromium d-orbital interactions along the three Cartesian axis (x, y, and z).

ChemFord affords users the ability to instantiate interactable three-dimensional (3D) representations of the octahedral coordination sphere of the chromium complex, in addition to the 5d-orbitals of the chromium metal atom, to direct peer discussion towards these three conceptual points.

Methods

This study was conducted throughout the academic year of 2021/2022 at the University of East Anglia (UK). The School of Chemistry is a dual-intensive (research and teaching) department. Our participant cohort for this research were second-year undergraduate students enrolled on a module of compulsory inorganic chemistry study. We employed a pre-test/post-test experimental design, in which students completed the following instrument prior to, and following, our ConcepTests within the PI session.

Peer instruction self-efficacy instrument (PISE)

The PISE is a 21-item instrument scored on a five-point Likert scale (Miller *et al.*, 2015). The PISE was developed based on the Sources of Self-Efficacy in Science Courses survey (SOSESC; Fencl and Scheel, 2004), and Bandura (1997). For the PISE survey responses collected throughout the course of this study, we calculated Cronbach's alpha values of 0.88 for the pre-test stage, and 0.90 for the post-test stage. This demonstrates very good internal consistency. For both alpha values calculated, the removal of adapted item 12: ''I get a sinking feeling when I think of trying to tackle difficult chemistry problems'', resulted in a higher alpha-if-deleted value.

The structure of our PI session is outlined in Fig. 3. Student response (voting) data for our six ConcepTests were collected through TurningPoint (2022), an audience response system in which students submitted their responses using mobile phones. In parallel, all students' PI discussions, alongside their interactions with ChemFord, were captured using audio- and screen-recording software installed on a suite of iPads distributed to student groups. This allowed the study of learning from two perspectives:

i. Probing the conceptual understanding of students through the collection of voting data.

ii. Studying the process of conceptual development during AR-supported peer discussion, through recorded conversations.

The research questions investigated were as follows:

Research question 1

How does the integration of augmented reality (AR) support students' PI discussions, and what types of interactions are occurring?

Research question 2

How does ConcepTest difficulty influence students' responses?

Research question 3

How does self-efficacy influence students' ConcepTest responses?

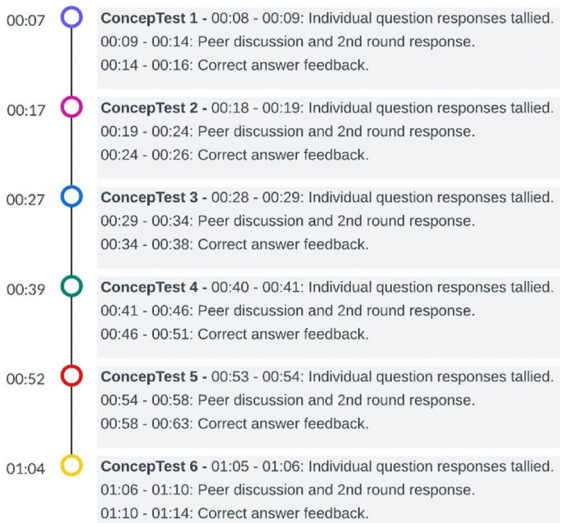


Fig. 3 A timeline of our PI session. Numbers preceeding each action indicate the session time in minutes.

Open Access Article. Published on 19 April 2024. Downloaded on 5/10/2026 12:06:01 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Ethical clearance was obtained under the regulations of UEA's School of Science Research Ethics Committee, a sub-committee of the UEA Research Ethics Committee. Participants were informed that their involvement within any aspect of this research was completely voluntary. In addition, Participants were made aware of their right to withdraw from the study, at any part of the research phase, without declaring a reason. Throughout the research period, participants were assured of data anonymity and confidentiality. Identifying information was irrevocably stripped from data documentation, and study codes utilized in their place. All information was stored securely and only accessible to the researcher.

Results

All collected data was transcribed verbatim, prior to being subjected to thematic analysis (using the approach of Braun and Clarke, 2006) for commonly occurring themes. For this analysis, themes were constructed around identified evidence of resource activation, as defined by Tuminaro and Redish (2007). The critical elements of this model are the basic elements of knowledge stored in long-term memory, the way those elements are linked, and the way in which those structures are activated in different circumstances (Tuminaro and Redish, 2007). The control systems that are considered here are epistemic games (Tuminaro and Redish, 2007).

Students are often unaware that they are engaging in a particular epistemic game. As such, the focus of this qualitative analysis is the interaction between students' AR experiences and the activation of these control structures. We start by examining ConcepTests 2 and 3, as these both showed significant intragroup improvement and high PI efficiency. ConcepTest 2 relates to the identification of a linear complex's crystal field splitting diagram, whereas ConcepTest 3 concerns the geometric [Jahn–Teller] distortion of a non-linear molecular system. These ConcepTests are examples of a productive dialogue in which a change in student thinking, and voting response, are evident.

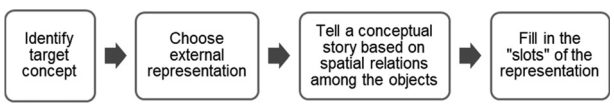


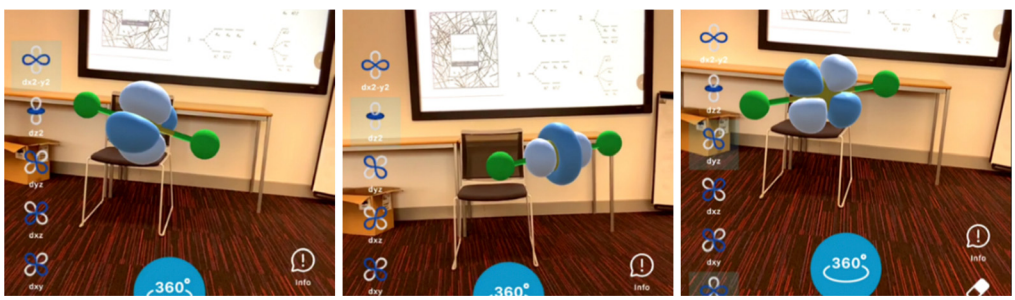
Fig. 4 Schematic diagram of some moves in the epistemic game Pictorial Analysis. Adapted from Tuminaro and Redish (2007).

Throughout discussions relating to ConcepTests 2 and 3, evidence of the Pictorial Analysis epistemic game was apparent (Fig. 4). In the Pictorial Analysis Game, students generate an external spatial representation that specifies the relationship between influences in a problem statement (Tuminaro and Redish, 2007). The epistemic form being a representation that the student generates to guide their inquiry.

The discussion presented (Table 1) was between a pair of students, of which one voted correctly on ConcepTest 2, and the other incorrectly. The second comment from group member (GM) 1 is the first activating statement in this dialogue. GM2 explains the interaction between the d-orbitals of the gold atom, and the two chlorido ligands. Amidst choosing a new representation on the virtual molecule, it is clear that there has been a change in thinking for GM1. This can be interpreted as an activating event, and evidence of the lowest level of resource activating (activation of a knowledge element). As the dialogue progresses, it is clear that GM1 has understood the concept, and is now able to use their knowledge to contribute to the discussion. Combining the video recording, representing the students' AR experience, with the audio recording of the peer discussion, gave a clear indication of the positive impact that using AR had on supporting students' thinking and knowledge construction.

As the session progressed to ConcepTest 3, employment of the Pictorial Analysis epistemic game was, again, evident from students' discussions. The example outlined in Table 2 is a group of three students, in which a single member answered

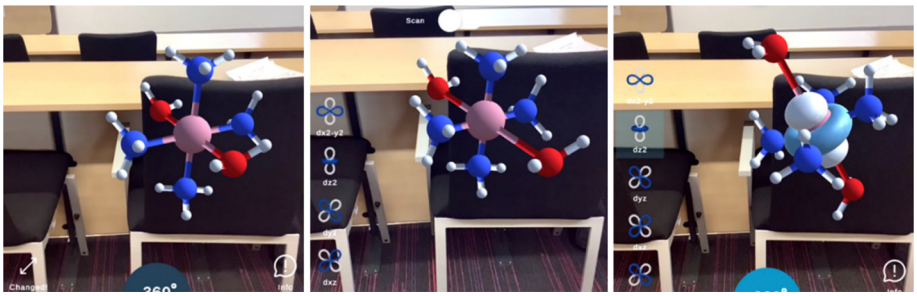
Table 1 Students' discussions for ConcepTest 2, showing some moves in the Pictorial Analysis epistemic game

Identify target concepts	**GM1:** "I think the answer is 2 [B], what do you think?"
	GM2: "Because it's linear, it has two point charges. In one axis, it will be higher in energy, and lower in energy in the other two."
Choose external representation	**GM1:** "So, this one would be low [$d_{x^2-y^2}$]?"
Tell a conceptual story based on spatial relations among the objects	**GM2:** "So, just the d_{z^2} would be higher in energy. It's pointing where the ligands are. This one [$d_{x^2-y^2}$] doesn't point towards where the ligands are."
Fill in the "slots" of the representation	**GM1:** "This one [d_{yz}] is in between, at the second level. The orbitals with x- and y-components are the lowest in energy."
	GM2: "They have the least interaction with anything. That's the way I understood it, so it is 2 [B]."
Corresponding AR experience	

 This journal is © The Royal Society of Chemistry 2024

Open Access Article. Published on 19 April 2024. Downloaded on 5/10/2026 12:06:01 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Table 2 Students' discussions for ConcepTest 3, showing some moves in the Pictorial Analysis epistemic game

Identify target concepts	**GM1:** "Let's look at the Jahn–Teller distortion." **GM3:** "So, these two ligands [*z*-axis] have moved away." **GM2:** "I put the answer as stabilising the *z*-components. That means the charges are further away, meaning it is lower energy."
Choose external representation	**GM1:** "That would make sense, wouldn't it?" **GM2:** "If you think that d_{z^2} was higher in energy before, it is now lower in energy. It is purely *z* component and is stabilised." **GM3:** "This is actually a good way of showing it. They are further apart [the ligands] so there is less repulsion."
Tell a conceptual story based on spatial relations among the objects	**GM1:** "I don't understand how you got the right answer." **GM2:** "Okay, so look at this molecule, and put on d_{z^2}. This is the situation before the distortion. Let's focus on d_{z^2}. With the distortion, the *z*-ligands get further away so the interaction is less. The *x*- and *y*-components experience greater interaction because they get closer."
Fill in the "slots" of the representation	**GM1:** "The distortion is stabilising the *z*-component. So, the answer is A."
Corresponding AR experience	

correctly during round 1, and the other two incorrectly. The interplay that is of particular interest is Section 2. GM2 is able to warrant proof of their claim through the use of ChemFord. The distortion of the represented octahedral complex is used as a means of activating the thinking of GM3. GM3 demonstrates activation of a knowledge structure, specifically, support knowledge building. The thinking of GM1 has not changed. Thus, GM2, building on their previous statement, thinks of a new way to persuade GM1 regarding the stabilisation of the *z*-components. Using ChemFord, GM2 is able to introduce the metal d-orbitals to support their conceptual story. Subsequently, GM1's thinking is activated, and repeats the statement that altered their perspective, reaching the correct conclusion. Both ConcepTests provide evidence of resource activation by means of successful AR-supported dialogue. All three students responded correctly on the second vote.

ConcepTest 1 is an interesting case. Although quantitative response data suggests that a majority of students answered correctly, qualitative data suggests that students may not have demonstrated a clear understanding at the start of the dialogue. As such, there are points of interest in terms of resource activation through utilisation of AR. Below, we present an example of a dialogue from a pair of students for ConcepTest 1. The AR representations employed are shown in Fig. 5. Both answered correctly before and after discussion:

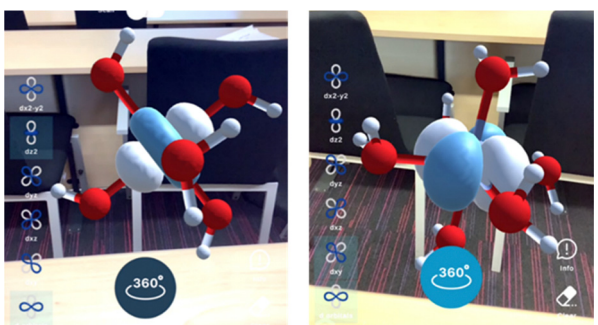


Fig. 5 AR representations employed during peer discussion of ConcepTest 1, with overlay of the d_{z^2} orbital (left), and the overlay of the d_{xy} orbital (right).

[Student group 4]	
GM1:	*"I put D; I don't know."*
GM2:	*"So, if we look at the d_{z^2} and the $d_{x^2-y^2}$, when it splits there will be 2 orbitals at the top and 3 on the bottom."*
GM1:	*"Yeah, but the question is why those ones?"*
GM2:	*"It's this [d_{z^2}] because the orbital is pointing towards the ligands. If you think of ligands as being point charges, the orbital overlaps with the ligands. That's higher energy. And this one also [$d_{x^2-y^2}$]. The ones between the axis are in the t_{2g}."*
GM1:	*"It makes sense that the top two orbitals are in line with the ligands, that these ones [$d_{x^2-y^2}$] are pointing towards the ligands which is unfavourable, so it's going to be the highest energy. These ones [d_{xy}] are between the axis and therefore lower in energy."*
GM2:	*"These aren't pointing at the ligands so I think that these are energetically favourable."*

Lastly, we provide an example from ConcepTest 5. Our data shows that this ConcepTest had the lowest correct response

Open Access Article. Published on 19 April 2024. Downloaded on 5/10/2026 12:06:01 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

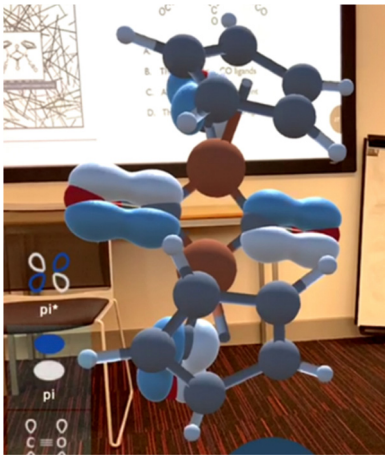


Fig. 6 3D representation of cyclopentadienyliron dicarbonyl dimer with superimposed carbonyl π bonding molecular orbitals from ConcepTest 5.

rate, as well as the lowest theoretical (and measured) PI efficiency. Furthermore, it was the only ConcepTest where the correct response rate of students was lower after discussion. Hence, it is important to understand the interactions present throughout discussions of ConcepTest 5, and how these differ from the successful dialogues presented in ConcepTests 1–3.

ConcepTest 5 asked students to use their understanding of pi backbonding to identify which carbonyl ligands (Fig. 6) are most susceptible to electrophilic attack. In all of the transcripts, a common theme was whether or not students could recognise that the two bridging carbonyl ligands are equivalent.

[Student group 19]
GM1: *"It's definitely not the top bridging CO."*
GM3: *"The top one will be sterically hindered by the two other ligands."*
GM2: *"Looking at the molecule you can see that both of the bridging COs are equivalent."*

For ConcepTest 5, we also provided representations of the π and π* molecular orbitals of the ligands, in addition to the iron atom d-orbitals in the hope of initiating discussion of electron backdonation. This was noted in some dialogues, in which students responded correctly during round 2:

[Student group 2]
GM1: *"I'm just thinking about the antibonding orbitals on the carbonyls. The antibonding for these ones [carbonyls] would be here."*
GM2: *"Okay."*
GM1: *"And these ones are bound to two metals and these ones are only bound to one metal."*
[Student group 5]
GM1: *"Yeah. Backbonding also provides electrons to the ligand."*
GM2: *"Yeah."*
GM1: *"So, those go into the pi* orbital of the CO."*

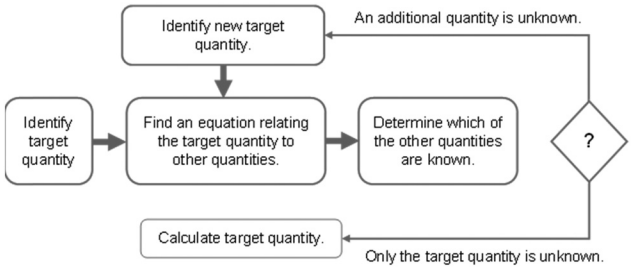


Fig. 7 Schematic diagram of some moves in the epistemic game Recursive Plug-and-Chug. Adapted from Tuminaro and Redish (2007).

Several dialogues for ConcepTest 5 provided examples of unproductive discussion, in which little conceptual chemistry was used. A reason for this may be that students were not able to retrieve the required knowledge elements to respond correctly, or that the AR experience did not manage to support resource activation. For group dialogues where the AR virtual objects were not referenced, or used as a driver for supporting the discussion, we found a greater number of incorrect responses after round 2. Evidence of the Recursive Plug-and-Chug epistemic game (Fig. 7) was also observed within ConcepTest 5 dialogues (not optimal). In the Recursive Plug-and-Chug epistemic game, students plug ideas into a problem situation and churn out answers without conceptually understanding the implications of their solution. Evidence of dialogue similar to that expected of Recursive Plug-and-Chug epistemic game was also observed in ConcepTests 4 and 6, but not in ConcepTests 1–3.

ConcepTest difficulty

For this study, we applied the analytical procedures of Item Response Theory (IRT; see Embretson and Reise, 2000) to calculate values of difficulty and discrimination for each ConcepTest. IRT models are non-linear monotonic functions describing the association between leaner ability on a latent variable and an item's characteristics on the probability of a particular response to that item (Embretson and Reise, 2000). A two-parameter model (2PL) was employed for our evaluation. The developed ConcepTests demonstrate reasonable difficulty and discrimination (Table 3).

We employed PI efficiency (η) calculations, defined with the help of Hake's standardised gain (Hake, 1998), to examine the effectiveness of each ConcepTest. The proportion of correct answers before, and after, the discussion is denoted by N_b and N_a respectively. While Hake's gain represents individual learning gain, PI efficiency is considered to reflect the ease of understanding gained through PI (Table 4). The collected response data from our ConcepTests was found to be normally distributed. Hence, we conducted paired-samples *t*-tests, alongside analysis of effect size, for intragroup comparisons. The theoretical value of N_a is expressed as a function of N_b (Nitta, 2010), with the theoretical value of $\eta = N_b$. For this study, the average difference between the measured, and theoretical values of $\eta = 0.061$, similar to a value of 0.062 recorded

 This journal is © The Royal Society of Chemistry 2024

Open Access Article. Published on 19 April 2024. Downloaded on 5/10/2026 12:06:01 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Table 3 IRT coefficients (2PL) for the six developed ConcepTests

ConcepTest No.	Difficulty	Discrimination
1	−2.000	0.636
2	−0.225	0.543
3	0.917	0.956
4	1.414	0.561
5	3.004	0.350
6	1.056	1.580

by (Nitta *et al.*, 2014) when measuring the effectiveness of PI using the Force Concept Inventory. The proportion of correct responses during independent voting in round 1 ranged from 0.290–0.897. The ideal range is reported to be from 0.35–0.70 (Crouch and Mazur, 2001). For ConcepTests 4–6, where correct independent response rates lie at the lower end of this range, students were likely to have had ineffective discussions during round 2. As such, the value of η observed is low.

The normalised proportion of correct responses before, and after, the discussion phase of each ConcepTest is shown in Fig. 8. We observed statistically significant improvement for correct response rates between the first and second round of voting on ConcepTests 2 and 3. For ConcepTests 1 and 4, this improvement was approaching significance, with the difference between groups greater than 0.2 standard deviations.

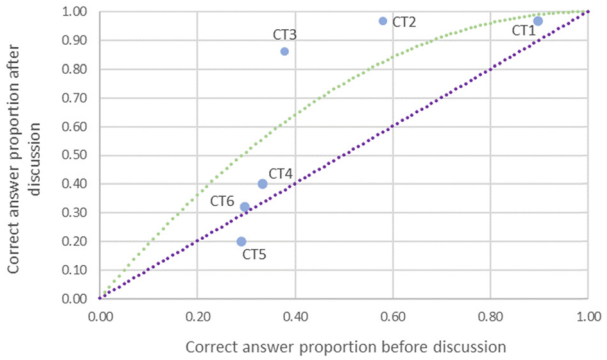


Fig. 8 Proportion of correct responses before, and after, discussion of each ConcepTest. The green line represents the theoretical curve for PI efficiency (Nitta, 2010). The purple line represents no change in correct pre-discussion and post-discussion response rate. Points above this line indicate improvements in accuracy, whereas points below the line represent decrements in accuracy.

Response switching

The descriptive statistics outlining the extent to which students switched their responses to each ConcepTest, between the first and second round of voting, are shown in Table 5. We provide details regarding the proportion of students who switched in any direction, in addition to the proportion of responses that are switched in a specific direction. Throughout our PI session, the results of round 1 voting were shared with the cohort prior to round 2. Previous work has shown that it is possible to bias the quality of peer discussion (and hence switching) by allowing students the opportunity to see response graphs (Perez *et al.*, 2010). However, when the results of an initial vote are evenly split between two or more answers, then displaying the graph may be a valuable conversation starter.

When switching is measured, it is important to ensure that the data is not confounded with the frequency of correct (or incorrect) responses in round 1 (Miller *et al.*, 2015). Normalising our response data with respect to students' answers in round 1 provides an adjusted measure of switching, independent of how many times a student was correct, or incorrect, in round 1.

Coupling these normalised values with the output of our 2PL IRT model allows us to examine switching as a function of ConcepTest difficulty (Fig. 9). A Pearson's correlation showed a strong, positive correlation, $r = 0.910$, between response switching and ConcepTest difficulty which was statistically significant, $p = 0.012$. With increasing ConcepTest difficulty, students are more likely to switch their answers from right-to-wrong ($r = 0.754$, $p = 0.084$), and wrong-to-different wrong ($r = 0.829$, $p = 0.042$). In addition, students are less likely to switch their answers from wrong-to-right ($r = -0.771$, $p = 0.072$). A finding consistent with previous studies (Miller *et al.*, 2015).

It is important for instructors to understand that they have some control over the measure of response switching that occurs throughout PI *via* the difficulty of the ConcepTests posed. Within our session, we attempted to scaffold this by posing easier ConcepTests first, subsequently building up to more difficult ConcepTests. Research has shown that prefacing more difficult problems with a sequence of related, but more

Table 4 Correct answer proportion and PI efficiency of our CTs

	ConcepTest					
	1	2	3	4	5	6
No. of respondents before PI discussion	29	31	29	33	31	27
No. of respondents after PI discussion	32	31	29	25	30	25
Correct answers before discussion (N_b)	0.897	0.581	0.379	0.333	0.290	0.296
Correct answers after discussion (N_a)	0.969	0.968	0.862	0.400	0.200	0.320
Paired Samples *t*-test	0.161	<0.01	<0.01	0.161	0.375	1.000
Cohen's *d**	0.28	0.77	0.91	0.29	0.17	0.00
Theoretical value of N_a	0.989	0.824	0.614	0.555	0.496	0.504
PI efficiency (η)	0.699	0.924	0.778	0.100	−0.127	0.034
Theoretical value of PI efficiency (η)	0.897	0.581	0.379	0.333	0.290	0.296
Difference between theoretical and measured values	0.198	−0.343	−0.399	0.233	0.417	0.262

Open Access Article. Published on 19 April 2024. Downloaded on 5/10/2026 12:06:01 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Table 5 The proportion of students' responses that were switched between the first and second round of voting. [Wrong-to-right (W–R); wrong-to-wrong (W–W); and right-to-wrong (R–W)]

ConcepTest	Students who switched (%)	Direction		
		W–R	W–W	R–W
1	29.41	70.0	10.0	20.0
2	48.48	81.3	12.5	6.3
3	65.63	71.4	23.8	4.8
4	63.64	28.6	38.1	33.3
5	65.63	19.0	47.6	33.3
6	54.84	23.5	52.9	23.5

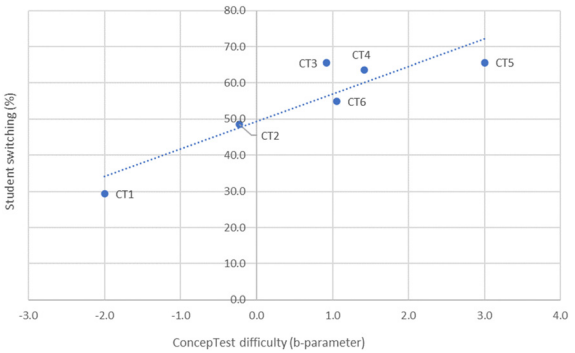


Fig. 9 Student switching (%) in any direction for each ConcepTest as a function of difficulty. Each point represents a different ConcepTest.

basic conceptual questions, helps students answer harder problems (Ding *et al.*, 2011). Cognitively, presenting easier questions prior to difficult questions may help students break down concepts into smaller, more manageable chunks when questioning the same concept. As ConcepTests often require students to apply conceptual understanding in new contexts, it is possible that scaffolding difficult ConcepTests may assist with positive switching transitions. A future study of ConcepTest response patterns to a series of scaffolded discussion points would prove interesting in providing further insight into the relationship between switching and ConcepTest difficulty.

Reported self-efficacy

From examination of students reported self-efficacy measures, we find evidence of a relationship between response switching and pre-session self-efficacy. Students returning lower measures of self-efficacy were more likely to switch their responses negatively in comparison to students with higher self-efficacy. Additionally, students reporting higher self-efficacy were more likely to switch from *wrong-to-right*. Fig. 10 shows the normalised proportion of switched responses for students with lower and higher self-efficacy. To analyse differences in self-efficacy, we divided the cohort into the top and bottom 27% (Preacher, 2015).

Students with higher pre-session self-efficacy negatively switched less often, and positively switched more often than students with low self-efficacy. As we did not administer a pre-test assessment, we are unable to control for covariates such as

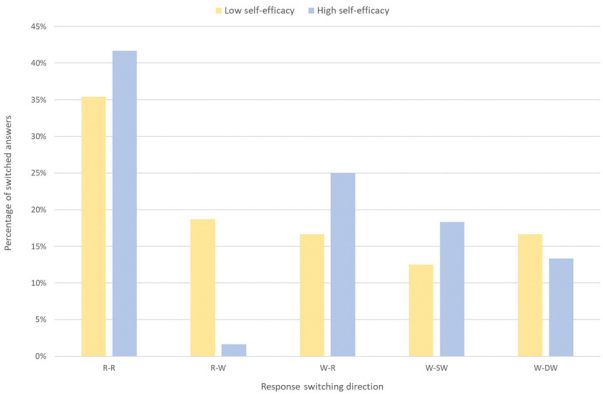


Fig. 10 Response switching patterns for students in the top and bottom 27% of reported self-efficacy measures.

prior knowledge, but previous work has indicated that self-efficacy may be more predictive of switching than incoming knowledge (Zajacova *et al.*, 2005). In addition, students with high self-efficacy positively switched their responses more often than students with low self-efficacy.

Students' responses to two individual items on the PISE moderately correlated with not switching responses. These statements are: *"I usually don't worry about my ability to solve chemistry problems"*, ($p < 0.01$); and *"I know how to explain my answers to organometallic chemistry questions in a way that helps others understand my answer"*, ($p < 0.01$). In contrast, these two same items strongly negatively correlated ($p < 0.01$) with negatively switching responses. For item 10, students who either disagreed or strongly disagreed negatively switched significantly more than students who agreed or strongly agreed ($p < 0.001$). This difference was also observed for item 16 ($p = 0.01$). Students with a low assessment of their problem solving and science communication abilities are significantly more likely to negatively switch their responses than students with a high assessment of those abilities.

Following our PI session, median Likert scores on the PISE instrument improved on the following items: *"When I come across a tough chemistry problem, I work at it until I solve it"* (neutral to agree); *"I like hearing about questions that other students have about chemistry"* (neutral to agree); *"I can communicate science effectively"* (neutral to agree, $p = 0.04$); and *"I can communicate chemistry effectively"* (neutral to agree, $p = 0.025$).

Study limitations

The limitations of this study must be observed. Firstly, our PI evaluation is based on data gathered from one session and six ConcepTests. It is therefore difficulty to generalise the results. Quantitative data concerning conceptual understanding was collected solely through ConcepTest voting data. Objective tests, such as the Force Concept Inventory, have been used previously to evaluate entire PI sessions (Hestenes *et al.*, 1992). In addition, the data analysis was based upon a relatively small sample size. This was the result of modest enrolment for our PI session.

 This journal is © The Royal Society of Chemistry 2024

Open Access Article. Published on 19 April 2024. Downloaded on 5/10/2026 12:06:01 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Conclusions

This study contributes to the growing body of evidence on how students can benefit from the pedagogical approach of PI supported by AR. In summary, we have found evidence of the interaction between ConcepTest difficulty and response switching. It is important for instructors to understand that they have some control over the measure of response switching that occurs throughout PI *via* the difficulty of the ConcepTests posed. The difficulty values for our ConcepTests were calculated using a 2PL IRT model. The output of our IRT model showed adequate difficulty and discrimination values for our ConcepTests. We also evaluated the effectiveness of each ConcepTest using peer efficiency calculations. ConcepTests 1–3 demonstrated the highest values of PI efficiency.

Moreover, we examined the relationship between response switching and reported self-efficacy. Students reporting higher measures of self-efficacy displayed lower levels of switching in a negative direction. Students with a low assessment of their problem solving and science communication abilities were significantly more likely to switch their responses from right to wrong than students with a high assessment of those abilities. Through qualitative analysis, we have provided evidence of Pictorial Analysis through AR-supported PI discussions. Where calculated peer efficiency values for ConcepTests were lower, this was less apparent, with Recursive Plug-and-Chug being the commonly observed control structure. It would be interesting to see what would happen if they were asked the same questions again at a later date – retention. Or a third round of the same question to see how many make 3 different choices.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The study team would like to express thanks to all the students who participated in this study. We are grateful to Dr Anna K. Wood for helpful discussions prior to planning these activities.

Notes and references

Andrew S., (1998), Self-efficacy as a predictor of academic performance in science, *J. Adv. Nursing*, **27**(3), 596–603.

Bandura A., (1977), Self-efficacy: toward a unifying theory of behavioral change, *Psychol. Rev.*, **84**(2), 191–215.

Bandura A., (1986), *Social Foundations of Thought and Action*, Prentice Hall.

Bandura A., (1997), *Self-Efficacy: The Exercise of Control*, W. H. Freeman.

Bandura A., (2001), Social Cognitive Theory: An Agentic Perspective. *Annu. Rev. Psychol.*, **52**(1), 1–26.

Bartimote-Aufflick K., Bridgeman A., Walker R., Sharma M. and Smith L., (2015), The study, evaluation, and improvement of university student self-efficacy, *Studies Higher Educ.*, **41**(11), 1918–1942.

Boyer D. A., Zollo J. S., Thompson C. M., Vancouver J. B., Shewring K. and Sims E., (2000), A quantitative review of the effects of manipulated self-efficacy on performance, Poster session presented at the annual meeting of the American Psychological Society, Miami, FL.

Braun V. and Clarke V., (2006), Using thematic analysis in psychology, *Qualitative Res. Psychol.*, **3**(2), 77–101.

Crouch C. and Mazur E., (2001), Peer Instruction: ten years of experience and results, *Am. J. Phys.*, **69**(9), 970–977.

Ding L., Reay N., Lee A. and Bao L., (2011), Exploring the role of conceptual scaffolding in solving synthesis problems, *Phys. Rev. Spec. Top.--Phys. Educ. Res.*, **7**(2), 020109.

Dogan U., (2015), Student Engagement, Academic Self-efficacy, and Academic Motivation as Predictors of Academic Performance, *The Anthropologist*, **20**(3), 553–561.

Elford D., Lancaster S. J. and Jones G. A., (2022), Exploring the effect of augmented reality on cognitive load, attitude, spatial ability, and stereochemical perception, *J. Sci. Educ. Technol.*, **31**(3), 322–339.

Elford D., Lancaster S. J. and Jones G. A., (2023), Augmented reality and worked examples: Targeting organic chemistry competence, *Computers & Education: X Reality*, **2**, 100021.

Embretson S. and Reise S., (2000), *Item Response Theory*, Psychology Press.

Fencl H. S. and Scheel K. R., (2004), Pedagogical approaches, contextual variables, and the development of student self-efficacy in undergraduate physics courses, *AIP Conf. Proc.*, **720**, 173–176.

Hake R., (1998), Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses, *Am. J. Phys.*, **66**(1), 64–74.

Halpern D. and Hakel M., (2003), Applying the Science of Learning to the University and Beyond: Teaching for Long-Term Retention and Transfer, *Change: The Magazine of Higher Learning*, **35**(4), 36–41.

Hestenes D., Wells M. and Swackhamer G., (1992), Force concept inventory, *The physics teacher*, **30**(3), 141–158.

Knight J. K., Wise S. B. and Southard K. M., (2013), Understanding clicker discussions: student reasoning and the impact of instructional cues, *CBE—Life Sci. Educ.*, **12**, 645–654.

Lancaster S. J., Cook D. F. and Massingberd-Mundy W. J., (2019), Peer instruction as a flexible, scalable, active learning approach in higher education, in Seery M. K. and Mc Donnell C. (ed.), *Teaching Chemistry in Higher Education: A Festschrift in Honour of Professor Tina Overton*, Creathach Press, Dublin, pp. 89–104.

Mazur E., (1997), Peer instruction: getting students to think in class, *AIP Conf. Proc.*, 981–988.

McConnell D., Steer D. and Owens K., (2003), Assessment and Active Learning Strategies for Introductory Geology Courses, *J. Geosci. Educ.*, **51**(2), 205–216.

Milgram P., Takemura H., Utsumi A. and Kishino F., (1995), Augmented reality: a class of displays on the reality-virtuality continuum, *SPIE Proc.*, 282–292.

Open Access Article. Published on 19 April 2024. Downloaded on 5/10/2026 12:06:01 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Miller K., Schell J., Ho A., Lukoff B. and Mazur E., (2015), Response switching and self-efficacy in Peer Instruction classrooms, *Phys. Rev. Spec. Top.--Phys. Educ. Res.*, **11**(1), 010104.

Moritz S., Feltz D., Fahrbach K. and Mack D., (2000), The Relation of Self-Efficacy Measures to Sport Performance: A Meta-Analytic Review. *Res. Quarterly Exercise Sport*, **71**(3), 280–294.

Newbury P., (2013), *Writing good peer instruction questions*.

Nitta H., (2010), Mathematical theory of peer-instruction dynamics. *Phys. Rev. Spec. Top.--Phys. Educ. Res.*, **6**(2), 020105.

Nitta H., Matsuura S. and Kudo T., (2014), Implementation and analysis of peer-instruction in introductory physics lectures, *J. Sci. Educ. Jpn.*, **38**, 12–19.

Perez K., Strauss E., Downey N., Galbraith A., Jeanne R. and Cooper S., (2010), Does Displaying the Class Results Affect Student Discussion during Peer Instruction? *CBE—Life Sci. Educ.*, **9**(2), 133–140.

Pietsch J., Walker R. and Chapman E., (2003), The relationship among self-concept, self-efficacy, and performance in mathematics during secondary school, *J. Educ. Psychol.*, **95**(3), 589–603.

Porter L., Bailey Lee C., Simon B. and Zingaro D., (2011), Peer instruction: Do students really learn from peer discussion in computing? in *Proceedings of the seventh international workshop on Computing education research*, pp. 45–52.

Preacher K., (2015), Extreme Groups Designs, *Encycl. Clin. Psychol.*, 1–4.

Rao S. P. and DiCarlo S. E., (2000), Peer instruction improves performance on quizzes, *Adv. Physiol. Educ.*, **24**(1), 51–55.

Ravna O. V., Garcia J., Themeli C. and Prasolova-Førland E., (2022), Supporting Peer-Learning with Augmented Reality in Neuroscience and Medical Education, in *KES International Conference on Smart Education and E-Learning*, Singapore: Springer Nature Singapore, pp. 299–308.

Schell J. and Mazur E., (2015), Flipping the chemistry classroom with peer instruction, *Chemistry education: Best practices, opportunities and trends*, pp. 319–344.

Schunk D., (2005), Self-Regulated Learning: The Educational Legacy of Paul R. Pintrich, *Educ. Psychol.*, **40**(2), 85–94.

Schunk D. H. and DiBenedetto M. K., (2016), Self-Efficacy theory in Education, *In Handbook of Motivation at School*, 2nd edn, pp. 34–52.

Schunk D. and Usher E., (2019), Social Cognitive Theory and Motivation, *Oxford Handbook Human Motivation*, 9–26.

Simon B., Kohanfars M., Lee J., Tamayo K. and Cutts Q., (2010), Experience report: peer instruction in introductory computing, in *Proceedings of the 41st ACM technical symposium on Computer science education*, pp. 341–345.

Smith M. K., Wood W. B., Adams W. K., Wieman C., Knight J. K., Guild N. and Su T. T., (2009), Why peer discussion improves student performance on in-class concept questions, *Science*, **323**, 122–124.

Stajkovic A., Lee D. and Nyberg A., (2009), Collective efficacy, group potency, and group performance: meta-analyses of their relationships, and test of a mediation model, *J. Appl. Psychol.*, **94**(3), 814–828.

Themelis C., (2022), Combining augmented reality with Peer Learning Pedagogy: IPEAR theoretical framework, AACE, available at: **https://aace.org/review/combining-augmented-reality-with-peer-learning-pedagogy-ipeartheoretical-framework/** (Accessed: 16 April 2024).

Tuminaro J. and Redish E. F., (2007), Elements of a cognitive model of physics problem solving: Epistemic games, *Phys. Rev. Spec. Top.—Phys. Educ. Res.*, **3**(2), 020101.

Vickrey T., Rosploch K., Rahmanian R., Pilarz M. and Stains M., (2015), Research-Based Implementation of Peer Instruction: A Literature Review, *CBE—Life Sci. Educ.*, **14**(1), es3.

Zajacova A., Lynch S. and Espenshade T., (2005), Self-Efficacy, Stress, and Academic Success in College, *Res. Higher Educ.*, **46**(6), 677–706.

 This journal is © The Royal Society of Chemistry 2024