Educación Química

Peer instruction in chemistry classes: systematic review on contributions and possibilities

Peer Instruction en clases de química: revisión sistemática sobre contribuciones y posibilidades

Cassiana Herzer Griebeler, Maurícius Selvero Pazinato y Camila Greff Passos[1]

Resumen

El objetivo de este trabajo es mapear estudios sobre Instrucción entre Pares en la Enseñanza de la Química. Para esto, desarrollamos una revisión sistemática de cómo se ha implementado e investigado la ***Peer Instruction*** (Instrucción entre Pares- IP) en clases de Química, así como el impacto de la IP en los resultados de aprendizaje de los estudiantes en ese contexto. Nuestro objetivo es apoyar a los profesores de química en la elección de utilizar la IP y promover su implementación basada en evidencia. La revisión abarca el período desde 1991 hasta 2022. Los estudios se llevaron a cabo principalmente en instituciones de educación superior, en universidades públicas de Inglaterra, Turquía, Estados Unidos y Brasil. La IP se implementó en cursos de Química, como en los subcampos de Química General, Química Física y Química Analítica. Algunos estudios adaptaron la IP a cursos en línea, mientras que otros compararon diferentes tipos de grupos de pares e investigaron la influencia de diferentes factores en la IP, como la lectura previa a la clase y la presentación de los resultados de votación en vivo. La efectividad de la IP varió según la forma en que se implementó, como la agrupación de pares, clases en línea o presenciales. Los resultados de la IP en las clases de química se han monitorizado principalmente a través de la percepción de los estudiantes, y el rendimiento en preguntas conceptuales se utiliza con frecuencia como herramienta de evaluación. Esta revisión encontró un pequeño número de estudios sobre la IP en química, especialmente en clases de química de secundaria. Además, las actividades previas a la clase no se emplean ampliamente. La mayoría de los artículos utilizaron análisis cuantitativos e indican resultados positivos para el proceso de aprendizaje.

Palabras clave: metodología activa; educación química; enseñanza de la química; métodos interactivos.

Abstract

This work provides a mapping of studies concerning Peer Instruction (PI) in the teaching of chemistry. To this end, a systematic review was undertaken of how PI has been implemented and investigated in chemistry classes, as well as the impact of PI on the learning outcomes of students. The review covers the period from 1991 to 2022. The reported studies were mainly conducted at higher education institutions, at public universities in England, Turkey, the USA, and Brazil. In chemistry courses, the PI method was implemented in subareas such as general chemistry, physical chemistry, and analytical chemistry. Some studies adapted PI to online courses, while others compared different types of peer groups and investigated the influence of different factors on PI, such as pre-class reading and showing live voting results. The effectiveness of PI varied depending on the way it was implemented, such as in peer grouping, online activities, or presential classes. The outcomes of PI in chemistry classes have mainly been monitored by means of the perceptions of the students, with performance in answering conceptual questions frequently being used as an assessment tool. This review found only a few studies concerning PI in chemistry, particularly in high school chemistry classes. Additionally, pre-class activities were not widely employed. Most of the articles utilized quantitative analysis and indicated positive outcomes for the learning process.

Keywords : active methodology; chemical education; chemistry teaching; interactive methods.

CÓMO CITAR:
Herzer Griebeler, C., Selvero Pazinato, M., y Greff Passos, C. (2024). Peer instruction in chemistry classes: systematic review on contributions and possibilities. ***Educación Química, 35***(4). https://doi.org/10.22201/fq.18708404e.2024.4.86704

---

[1] Universidade Federal do Rio Grande do Sul.

Introduction and background

Active learning has been frequently praised among teachers, instructors, and professors at educational institutions. The movement towards this mode is more recent in higher education, where traditional lectures are well-established (Pollozi et al., 2019). However, many studies have highlighted the benefits of greater engagement of students in their learning processes, leading to attention being paid to new appropriate pedagogical approaches (Børte, Nesje, & Lillejord, 2019; Freeman, 2014). Active learning can be considered an umbrella concept that includes a vast number of methodologies with varying levels of student engagement. It is a process whereby students construct meaning by actively engaging in the process of learning, rather than passively receiving information (Prince, 2004).

Difficulties reported by higher education chemistry instructors and lecturers, when using transmissive teaching methods, include inadequate conceptual understanding, reflected in low grades, especially in introductory courses (Nakhleh, 1992), and poor attitudes of students towards science (Simpson & Oliver, 1990). These problems can often be resolved by implementing evidence-based active teaching and learning pedagogies, such as Peer Instruction.

Peer Instruction (PI) is an active student-centered pedagogy, developed in 1991 by Eric Mazur, a physicist and educator at Harvard University (Mazur, 1999).

PI can essentially be described as follows:

1) Rapid revision of fundamental concepts of the topic;

2) Posing a conceptual question;

3) Allowing the students time to think;

4) The students answer individually;

5) The instructor checks the distribution of the replies and determines whether the students should proceed to the stage of discussion among peers (if the distribution of correct responses is between 35 and 70%), or advance to step 7;

6) Peer discussion about the students' answers;

7) The students provide answers again after revision;

8) The instructor checks and presents the answers to the students, and explains the question.

After a short presentation (approximately 20 minutes) about a fundamental concept covered in a class, the instructor presents a conceptual question (ConcepTest) to the students. The students are allowed 2 or 3 minutes to think, after which they select their answer using apps or flash cards. The instructor observes the distribution of answers and requests the students to convince their colleagues about their choices, if the distribution of correct answers is between 35 and 70%, or proceeds to the step of explaining the question and restarts the cycle by presenting a new question or moving on to the next topic (Mazur, 1999).

The PI pedagogy has been highlighted internationally due to its ability to actively engage students in the educational process. The methodology was developed as a response to a problem identified by Mazur. Using standardized tests, Mazur perceived that his students in the Introductory Physics course at Harvard University were able to correctly answer difficult quantitative questions, but not easy qualitative questions, indicating that the students memorized algorithms for solving quantitative problems, but failed to learn related concepts. Consequently, he conceived PI to improve the learning of fundamental concepts in the Introductory Physics course (Mazur, 1999).

Other studies have found PI to be particularly effective in enhancing conceptual understanding (Müller et al., 2017; Vickrey et al., 2015). This is one of the characteristics of PI that differentiates it from other active methodologies; it is structured specifically to stimulate conceptual development (Mazur, 1999). To a lesser extent, other work has also examined the use of PI to develop abilities such as self-efficacy and metacognition (Müller et al., 2017). In this way, PI differs from other active methodologies or approaches, including problem-based learning, where the main objectives are the development of higher order abilities for argumentation and persuasion, among others (Bernardi & Pazinato, 2022).

PI is a well-studied methodology in Physics Education and has been adopted in many other disciplines (Crouch & Mazur, 2001). Since its creation, it has been applied and investigated in various courses, including small and large classes, and in different contexts (Crouch & Mazur, 2001). Müller et al. (2017) conducted an extensive review on the use of PI, using two international databases: the Social Sciences Citation Index (SSCI) (via the Web of Science platform) and the Education Resources Information Center (ERIC). The authors found that most of the research was carried out at academic institutions in the United States, with a focus on STEM (Science, Technology, Engineering, and Mathematics) disciplines, particularly in the field of Physics.

The use of the methodology had beneficial effects on the students' understanding of concepts, problem-solving abilities, and overall academic performance. Additionally, this method fosters a positive attitude towards learning content and the pedagogical approach employed. Instructors often adapt and customize the implementation of inquiry-based teaching, seamlessly integrating it with other educational strategies, highlighting its adaptability and flexibility (Müller et al., 2017).

Frequently, adaptations are tailored to specific contexts regarding class size, technological tools, or the course itself, which indicates the flexibility of the method and its potential for integration with other approaches. Nevertheless, the literature must support such changes, to ensure effectiveness (Müller et al., 2017; Olpak & Yilmaz, 2022). However, there are relatively few studies about PI and its modifications in chemistry education, with most of them focusing on higher education (Müller et al., 2017; Olpak & Yilmaz, 2022). Müller et al. (2017) identified only two articles on the use of PI in chemistry education from 1991 to 2015. Olpak & Yilmaz (2022) reported on seven studies of this nature from 1997 to 2020.

Therefore, the aim of this work was to perform a mapping of studies concerning Peer Instruction in the teaching of Chemistry. For this, a systematic review was undertaken of how PI has been implemented and investigated in chemistry classes, as well as the impact of PI on the learning outcomes of students. The findings expand upon the evidence presented in review articles discussing the current state of PI in the STEM disciplines. For this purpose, three research questions were proposed and are addressed in this systematic review, which constitutes part of a PhD dissertation:

1. In what context has PI been implemented and investigated in chemistry classes?
2. What approaches and modifications have been adopted with the implementation of PI in chemistry classes? What occurs before, during, and after the class?
3. What aspects are being monitored, and how have they been evaluated? (a) What feedback do students provide regarding the implementation of this approach? (b) Is there evidence demonstrating that the approach leads to improvements in knowledge, attributes, and/or skills?

Methodology

To address the above questions, a systematic search was performed using the Web of Science (WoS), Education Resources Information Center (ERIC), and Scopus electronic databases. The criteria established for the results were (1) peer-reviewed articles and (2) papers that investigated and implemented PI in higher education or high school chemistry classes, following the definition provided above. This review covers articles published between 1991 (the year in which PI was created) and 2022, to identify the most recent trends concerning PI. After searching using the term "Peer Instruction AND chemistry", 220 articles were reviewed for relevance, based on the title and abstract. Studies were excluded if the keywords were not found in it; the article was published before 1991; the terminology PI was used to refer to another methodology or only for citation; or the goal was to investigate the use of clickers in general and not PI itself. The ten remaining papers were read fully, critically analyzed, and were suitable for this systematic review (Figure 1).

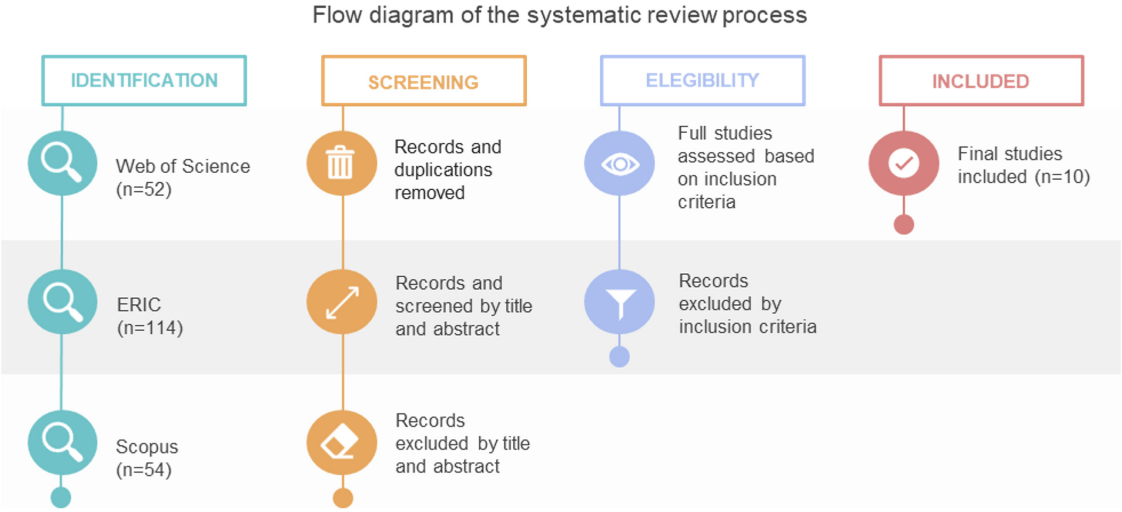

FIGURE 1. Flow diagram of the systematic review process.

After removing duplicates, the remaining articles were reviewed based on the abstracts, according to the inclusion/exclusion criteria. Finally, ten full texts (Aricò & Lancaster, 2018; Brooks & Koretsky, 2011; Bruck & Towns, 2009; Gok & Gok, 2016; Michinov, Morice, & Ferrières, 2015; Moraes, Carvalho, & Neves, 2016; Morice et al., 2015; Pearson, 2019; Belmonte, Borges, & Garcia, 2022; Yildirim & Canpolat, 2019) were found to be suitable for this systematic review. The articles were examined to extract and categorize data based on context, implementation (including pre-class strategies and in-class active learning interventions), monitored impacts, and methods of analysis. The data extracted were summarized and organized into codes related to the research questions, each containing its subcodes, as shown in Table 1. All ten articles were critically reviewed and were directly related to the questions proposed above. The main codes were determined *a priori*, as follows: (i) context; (ii) implementation; (iii) monitored impact and analysis of PI implementation. The subcodes were identified, revised, and refined using a combination of *a priori* and emergent coding.

TABLE 1. Operationalized codes and subcodes for context analysis.

Code	Subcode	Descriptors
Context	Education level	**HE**: Students sample at higher education. **BE:** Students sample at basic education.
	Major	Undergraduate major (e.g. Chemistry, Engineering, etc.) or discipline at high school (e.g. Chemistry).
	Course	**GC**: General Chemistry; **OC**: Organic Chemistry; **IC**: Inorganic Chemistry; **PC**: Physical Chemistry; **AC**: Analytical Chemistry; or grade at high school (10th grade).
	Sample	Size of the sample (participants).
	Status	**Pub**: public institution; **Priv**: private institute; **NI**: not informed in the article.
	Location	**NA**: North America; **EU**: Europe; **EU/AS**: Europe and Turkey; **SA**: South America.
Implementation	Pre-class activities	Brief description of pre-class activities performed (if done).
	PI implementation	**Original**: Peer Instruction sequence implemented exactly as it was developed. **Altered sequence**: Modified sequence of Peer Instruction steps. Alteration in the order of steps, insertion of a new step, or step omission. For example, the students must write short explanations for their answers before and after the first and second tests. **Display mode**: The way that researchers collected the students' answers during a Peer Instruction session (e.g. electronically or using cards). **Management of groups**: *Regulated*: Peer discussion groups controlled for group size and composition. *Semi-regulated*: Peer discussion groups controlled for group size only. *Diversified*: Peer discussion groups not controlled for group size or composition. *Not informed*: No information provided about group characteristics.
Monitored impact and analysis of PI implementation	Monitored impact	**Perception of students**: Any survey applied to collect the students' perceptions about the methodology. **Performance in conceptual questions**: Performance in the first and second tests. **Conceptual improvement**: Students' conceptual improvement, confirmed by analysis of written explanations or interviews. **Performance in final tests**: Performance in final questionnaires or validated tests. **Performance in pre- and post-tests**: Performance before and after peer discussion polls. **Students' confidence in answers**: Students' confidence in each answer chosen during PI sessions.
	Data collection	Instruments used to collect data during the research (e.g. open answer questionnaires, conceptual questions, etc.).
	Method of analysis	Tools used to analyze the data collected (e.g. statistical analysis, percentual analysis, etc.).

Results and Discussion

The ten studies that met the inclusion criteria were reviewed, and the outcomes were structured around three main codes, as described below. The criteria on which the codes were based are shown in Table 1. Chart 1 shows the distribution of articles over the years, for the period from 2009 to 2022, which was the period when publications were found that met the inclusion criteria of this review.

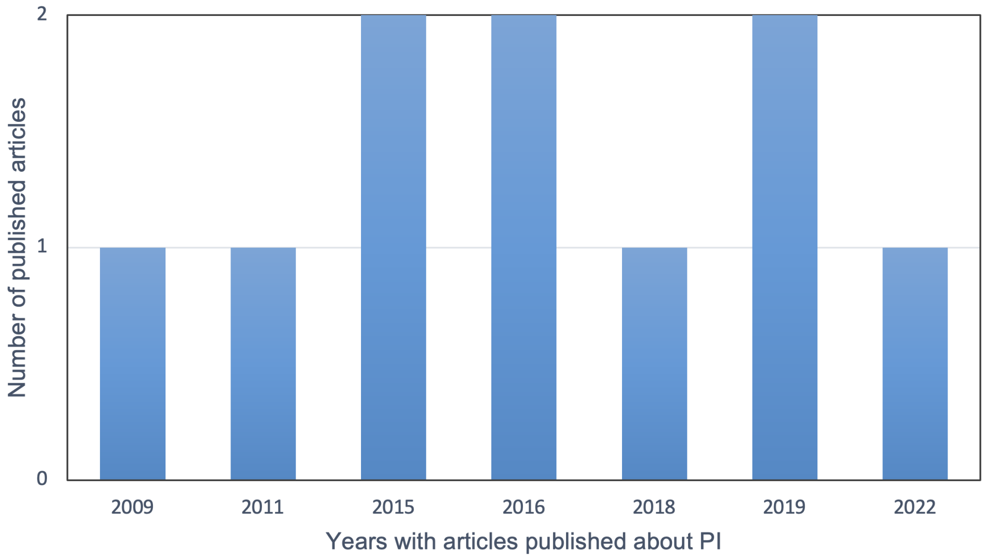

Chart 1. Distribution of published articles concerning PI, according to year, between 2009 and 2022.

Chart 1 only shows data for the years in which there were relevant publications (within the period studied, from 1990 to 2022). The results highlight the small number of articles concerning the use of PI in the teaching of chemistry, indicating the need for further research in this area, given the potential of PI evidenced in studies concerning the teaching of physics. (Zhang, Ding, & Mazur, 2017).

Context

The aim of the first question was to understand when and in what circumstances PI has been implemented in chemistry courses. Since Mazur's first paper and book publications on the methodology (Mazur, 1997, 1999), there have been many published studies concerning its application in various fields of education.

This systematic review, focusing on PI in chemistry education, identified the first article published in 2009, featuring the use of PI for teaching general chemistry (Bruck & Towns, 2009). Most of the studies (eight out of ten) were conducted in higher education settings. It was only in 2015, 2016, and 2019 that more than one article was published per year (Table 2, Chart 1)

Article	Ed. level	Major	Course	Status	Loc	Sample
Bruck & Towns (2009)	HE	Engineering, Science, Pharmacy	GC	Priv	NA	1100
Brooks & Koretsky (2011)	HE	Chemistry, Biology, and Environmental Engineering	PC	Pub	NA	128
Michinov, Morice, & Ferrières (2015)	HE	Chemical Engineering	AC	NI	NI	84
Morice et al. (2015)	HE	Chemical Engineering	AC	NI	NI	50
Gok & Gok (2016)	HE	NI	GC	Pub	EU/AS	47
Moraes, Carvalho, & Neves (2016)	BE	Chemistry	10th gr.	Pub	SA	163
Aricò & Lancaster (2018)	HE	NI	IC	Pub	EU	NI
Pearson (2019)	HE	Pharmacy	OC	NI	NI	223
Yildirim & Canpolat (2019)	BE	Chemistry	11th gr.	Pub	EU/AS	64
Belmonte, Borges, & Garcia (2022)	HE	Physical Chemistry, Chemical Engineering, Physical Engineering	PC	Pub	SA	54

HE: higher education; BE: basic education; NI: not informed; Pub: public university; Priv: private university; Loc: location.

TABLE 2. Codes for description of the contexts in which PI was implemented in the studies analyzed in this review.

The chemistry courses in which PI was implemented covered a range of subareas, with two studies reporting on general chemistry (GC) (Bruck & Towns, 2009; Gok & Gok, 2016), two on physical chemistry (PC) (Brooks & Koretsky, 2011; Belmonte, Borges, & Garcia, 2022), two on analytical chemistry (AC) (Michinov, Morice, & Fèrrieres, 2015; Morice et al., 2015), one on organic chemistry (OC) (Pearson, 2019), and one on inorganic chemistry (IC) (Aricò & Lancaster, 2018). The courses were mainly taken by students majoring in chemistry (Gok & Gok, 2016; Belmont, Borges, & Garcia, 2022) or chemical engineering (Brooks & Koretsky, 2011; Michinov, Morice, & Fèrrieres, 2015; Morice et al., 2015; Belmont, Borges, & Garcia, 2022), while only two papers investigated the implementation of PI in chemistry courses for alternative majors such as physical engineering (Belmont, Borges, & Garcia, 2022) and pharmaceutical sciences (Pearson, 2019). The other two studies (Moraes, Carvalho, & Neves, 2016; Yildirim & Canpolat, 2019) were implemented at the high school level. One of them was conducted in a 10th grade class at a public high school in Brazil (Moraes, Carvalho, & Neves, 2016), where the topic studied was stoichiometry. The other study was conducted in an 11th grade class studying solutions, in a public school in Turkey (Yildirim & Canpolat, 2019).

In higher education institutions, PI was used for course contents including chromatography (Michinov, Morice, & Fèrrieres, 2015), thermodynamics (Brooks & Koretsky, 2011), chirality (Pearson, 2019), functional groups (Pearson, 2019), stoichiometry (Moraes, Carvalho, & Neves, 2016), solutions (Yildirim & Canpolat, 2019), kinetics (Belmont, Borges, & Garcia, 2022), inorganic chemistry (Aricò & Lancaster, 2018), and general chemistry (Bruck & Towns, 2009; Gok & Gok, 2016). It is important to point out that all the studies described the use of PI for teaching of *chemical contents*, with no articles being found that mentioned its application in a *thematic* context, in contrast to Case Study and Problem-Based-Learning methodologies (Bernardi & Pazinato, 2022). In addition, it is not surprising that most of the studies were undertaken at higher education institutions, since PI was created for and first applied in that context (Mazur, 2001). The findings of this review indicate that little effort has been made to implement PI in basic level chemistry education (Müller et al., 2017), as also observed for other active teaching and learning methodologies at that educational level. In their recent review, Bernardi & Pazinato (2022) found that the literature reports few initiatives for dissemination of the Case Study methodology in basic level chemistry education.

Of the eight studies conducted in higher education, six reported the institutions where PI was implemented, with five of them being public universities in England (Aricò & Lancaster, 2018), Turkey (Gok & Gok, 2016), the United States (Brooks & Koretsky, 2011; Bruck & Towns, 2009), and Brazil (Moraes, Carvalho, & Neves, 2016; Belmont, Borges, & Garcia, 2022). It is noteworthy that studies in Brazil were not reported in the reviews by Müller et al. (2017) and Olpak & Yilmaz (2022). In the present study, two articles were found, indicating that the PI method is spreading in the field of chemistry education in Brazil, although further development of research is still needed. The study by Brook & Towns (2009) was the only one stated to have been performed at a private university and could be highlighted for its large sample size of 1100 participants, while all the other studies had participant numbers ranging from 47 to 487.

Implementation

In the analysis of the selected articles, examination was made of the approaches and modifications used by researchers to investigate the effectiveness of the PI active methodology in chemistry education. Relevant information is summarized in Table 3, where the information is classified into pre-class and in-class activities (post-class activities were not used in any of the articles).

TABLE 3. Codes of pre-class and in-class activities, and adaptations of PI implementation made in the studies reviewed.

Reference	Pre-class activities	In-class activities		
		PI implementation	Display mode	Conceptual questions source
Bruck & Towns (2009)	-	Original	Electronic	Literature / by professors
Brooks & Koretsky (2011)	-	Altered sequence, just-ans*, assigned	Electronic	Literature / by professors
Morice et al. (2015)	-	Altered sequence, dsp-ans*, assigned	Electronic	By professors
Michinov, Morice, & Ferrières (2015)	-	Altered sequence	Electronic	Not informed
Gok & Gok (2016)	-	Original	Cards	By professors
Moraes, Carvalho, & Neves (2016)	Assigned reading and assignment quiz	Original	Electronic	By professors
Aricò & Lancaster (2018)	Watch recorded lectures, answer open questions, and write questions	Altered sequence, free-int*	Electronic	By professors
Pearson (2019)	-	Altered sequence, dsp-ans*	Electronic	By professors
Yildirim & Canpolat (2019)	Reading available material	Original, test*	Cards	By professors
Belmonte, Borges, & Garcia (2022)	Watch indicated videos or reading available material	Altered sequence	Electronic	Based on literature questions – by professors

*Free-int: free interaction between groups during peer discussion time; just-ans: justified answers; dsp-ans: answers displayed during peer discussion time; test: short test at the beginning of the class.

Of the ten studies included here four used pre-class activities as a preparation task for PI sessions (Aricò & Lancaster, 2018; Belmonte, Borges, & Garcia, 2022; Moraes, Carvalho, & Neves, 2016; Yildirim & Canpolat, 2019). This was a small number, since PI has been considered a good match with the flipped classroom methodology. In the study by Aricò & Lancaster (2018), the students were requested to view screencast recordings at home before each class, answer open questions, and write their questions on the same online platform. The instructor then accessed the questions and used them to construct the answers to conceptual questions used during PI classes (a Just-in-Time Teaching (JiTT) approach) (Novak et al., 1999). Belmonte, Borges, & Garcia (2022) also incorporated pre-class activities that required students to watch video lessons. However, they allowed the students the option to choose between watching the indicated videos or reading the available material. Yildirim & Canpolat (2019) and Moraes, Carvalho, & Neves (2016) asked the students to complete assigned reading tasks, while the latter also required the students to perform pre-class tests. The manual for PI use recommends pre-class preparation for PI sessions, since the classes can proceed more smoothly when the students have already been exposed to the material under study. On the other hand, when there is little or no pre-class student preparation, in-class activities become more difficult, due to the short time available for review of the content during PI sessions. Lucas (2012) reported that students who completed pre-class activities performed better in answering in-class conceptual questions and in the final grades. Despite this, most of the studies evaluated did not mention any activities related to pre-class preparation.

The reported in-class activities were directly related to the implementation of PI. Four of the studies (Bruck & Towns, 2009; Gok & Gok, 2016; Moraes, Carvalho, & Neves, 2016; Yildirim & Canpolat, 2019) applied PI in its original form, as established by its developer (Eric Mazur), which is denoted here as Original PI. Out of those studies, only two utilized pre-class reading activities to prepare students for PI sessions (Moraes, Carvalho, & Neves, 2016; Yildirim & Canpolat, 2019), with one of them (Yildirim & Canpolat, 2019) assigning grades to the pre-class activities. However, assigning grades may not be the most effective way to encourage students to complete the reading or video-watching before class, as pointed out by Heiner, Banet, & Wieman (2014). The remaining studies (Aricò & Lancaster, 2018; Brooks & Koretsky, 2011; Morice et al., 2015; Michinov, Morice, & Ferrières, 2015; Pearson, 2019; Belmonte, Borges, & Garcia, 2022) modified the sequence of steps in PI, so they were coded as Altered Sequence. For instance, one study (Brooks & Koretsky, 2011) added a brief written explanation to accompany the multiple-choice answers in the PI process, which allowed the authors to investigate in greater depth the influence of peer discussion on learning improvement. In the implementation by Aricò & Lancaster (2018), the students attended 50-minute PI sessions with 8-10 questions. For each question, a self-assessment rating, using a Likert scale, was employed to assess the students' confidence and mastery of skills. The authors also collaborated with the students in the creation of new conceptual questions for the subsequent PI sessions, using an open text option to collect the answers of the students, together with a platform that allowed them to construct their own questions or comment on others.

In another study, Morice et al. (2015) modified the PI guidelines and compared learning improvement with a control group where students did not share their thoughts with peers but were active (searching their notes and the internet) during the period between the first and second polls. In this case, all the students had a hybrid learning process, with online classes (8 hours) and a presential session (2 hours) at the end of the semester, in which both the experimental and control groups answered 14 conceptual questions in a row, without interspersing with short lectures. In this case, it should be noted that the students had live access to the choices of others during the first poll.

Pearson (2019) analyzed the influence on students' responses of displaying the answers live while the students discussed (in pairs) their choices. An extended time for peer discussion was also provided, to determine if it would alter the responses. In the work of Aricò & Lancaster (2018), conceptual questions were presented in various types of classes (lectures, workshops, and problem classes), all during one semester.

Michinov, Morice, & Ferrières (2015) compared procedures denoted "Stepladder Peer Instruction" and "Individual Peer Instruction" with "Classic Peer Instruction". In all three cases, the students firstly answered 10 multiple-choice questions in a row. Then, in the "Stepladder Peer Instruction" group, the participating students compared their answers and discussed them (in pairs) for 10 minutes. Two more participants joined the group, one at a time, at intervals of 10 minutes. At the end of 40 minutes, the students again responded individually to the 10 chromatography multiple-choice questions, presented in the same order as before. In the "Classic Peer Instruction" group, the process was the same, but all the participants of each group together discussed all 10 questions for 40 minutes. Additionally, in the "Individual Peer Instruction" group, the students did not share their reasoning. Instead, they used the extra 40 minutes to rethink their answers individually. The main issue the authors were trying to resolve was the uncertainty about the participation of all the students in the peer discussion during a PI session. This is a particular concern among academic researchers, since students show different levels of engagement in peer discussion (Crouch et al., 2007; Heiner, Banet, & Wieman, 2014), for reasons such as shyness, lack of confidence, poor understanding, or lack of motivation (Michinov, Morice, & Ferrières, 2015).

Belmonte, Borges, & Garcia (2022) adapted PI for an online course, in response to the COVID-19 pandemic, using the methodology to encourage student participation in remote classes. The authors proposed either a pre-class reading or an online video to prepare the students for asynchronous classes. PI occurred during synchronous online classes in which the students were assigned (in pairs) to separate online rooms, where they discussed various exercises for an hour and a half, after which they submitted their written answers to the instructor by means of an online platform.

Therefore, it is evident that alterations to the PI method have been made in various ways. According to the observations of Müller et al. (2017), the original PI method includes activities involving pre-class readings. However, due to the integration of PI with other methods such as JiTT, pre-class tasks such as readings are now considered part of JiTT, while those carried out in the classroom are attributed to PI. This may explain the absence of references to reading stages in the articles analyzed in this study.

The "Display mode" concerns the method used to collect the answers of the students during PI classes. There was a predominant use of electronic media to compile the answers, as observed in the review by Olpak & Yilmaz (2022), where most of the response technologies involved the use of clickers. Only two studies (Gok & Gok, 2016; Yildirim & Canpolat, 2019) used cards to display the responses. The primary justification for using cards is that it saves time, when compared to the use of electronic devices, and avoids the distractions that may be caused by the internet. On the other hand, cards make data collection difficult, while the students may be able see each other's answers, depending on the seating configuration and the number of students in the room. Table 2 also shows that although some instructors used the literature as the source for conceptual questions, most of them created the questions independently.

Information about the characteristics of peer discussion groups is also important in investigations concerning PI. The type and management of activities during this step of PI can influence the answers of the students, their confidence, and the quality of their engagement (Olpak & Yilmaz, 2022). It is known that fixed peer groups are associated with more positive beliefs and attitudes of students in PI classes (Zhang, Ding, & Mazur, 2017). In this review, half of the studies (Brooks & Koretsky, 2011; Morice, Michinov, & Ferrières, 2015; Morice et al., 2015; Pearson, 2019; Belmonte, Borges, & Garcia, 2022) were considered to use "regulated" groups (Figure 2), where the authors established the group size, preventing too many students entering the discussion and ensuring that all the students participated in it. The same group composition was also maintained for all conceptual questions (at least in the same class), so that the students could become more comfortable in presenting their arguments in response to the questions.

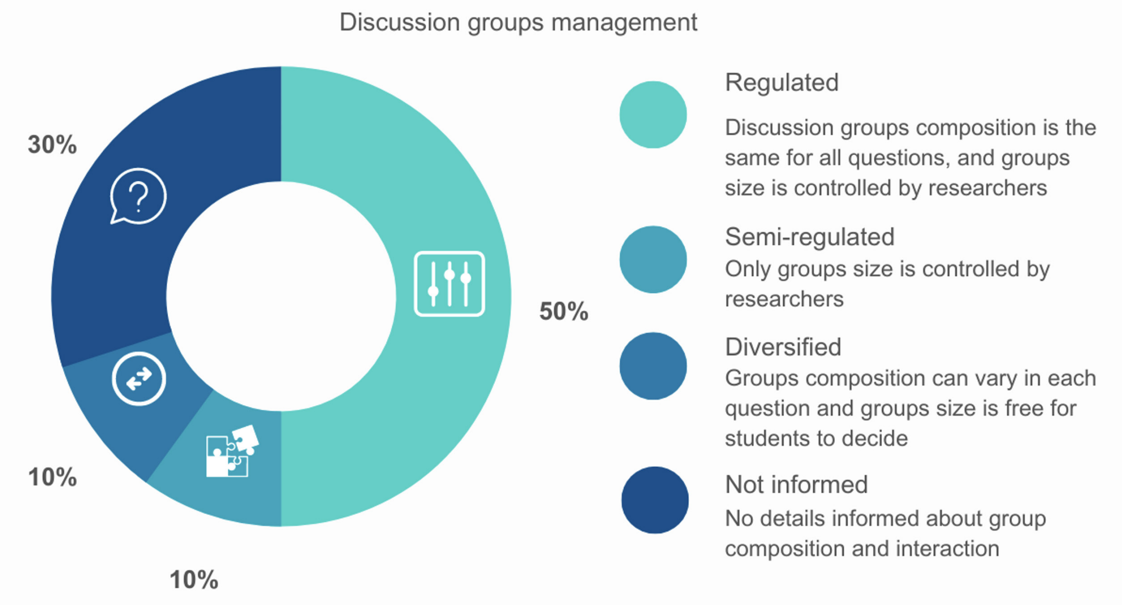

FIGURE 2. Management of discussion groups during implementation of PI in the studies analyzed in this review.

The "semi-regulated" category included the only study (Moraes, Carvalho, & Neves, 2016) that defined the group size, although no information was provided about maintaining the initial group structure. The "diversified" category included one study (Yildirim & Canpolat, 2019) that did not establish a group size, while the group composition could vary between questions. Three studies (Aricò & Lancaster, 2018; Bruck & Towns, 2009; Gok & Gok, 2016) did not provide specific information about group characteristics and were categorized as "no details informed". It is worth mentioning that four of the five studies that were categorized as "regulated" (Morice, Michinov, & Ferrières, 2015; Morice et al., 2015; Pearson, 2019; Belmonte, Borges, & Garcia, 2022) implemented an altered sequence of PI. It was likely that the researchers who used the original PI method may not have provided details about peer discussion groups ("no details informed") because they used the method as originally described by Eric Mazur or were not investigating aspects of group configuration during their research.

Monitored impact and analysis of PI implementation

The outcomes of PI implementation in the studies were monitored from several different perspectives, as shown in Figure 3.

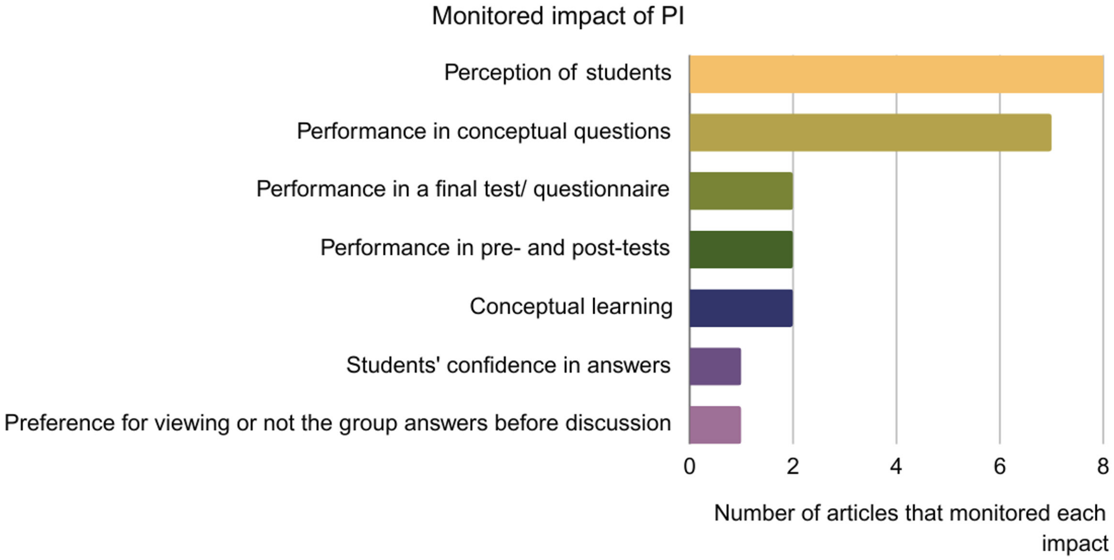

FIGURE 3. Monitored impacts of PI in the studies analyzed in this review.

It is known that the perceptions of students regarding the teaching pedagogy can influence their engagement in the learning process (Crouch & Mazur, 2001). Eight out of the ten studies included in this review used the students' perceptions as a measure of the outcome of the methodology, with data obtained using agreement scale and open-ended questionnaires (Aricò & Lancaster, 2018; Gok & Gok, 2016; Moraes, Carvalho, & Neves, 2016; Morice, Michinov, & Ferrières, 2015; Morice et al., 2015; Pearson, 2019; Belmonte, Borges, & Garcia, 2022; Yildirim & Canpolat, 2019). Aricò & Lancaster (2018) used a modified PI methodology throughout an entire descriptive inorganic chemistry course. The students were requested to watch recorded lectures at home, before each class, and attend 50-minute sessions of PI with 8-10 questions in a row. However, some students considered that too much time was spent watching recorded lectures and that more time was needed for explanation. Gok & Gok (2016) used a student evaluation questionnaire to assess the students' perception of the methodology, where the responses showed that students had a positive perception in terms of both affective and cognitive aspects of PI. They found it easy to follow, liked the interaction with peers and the opportunity to express their thoughts, and felt that PI helped in understanding the topics and in encouraging metacognitive skills.

Morice et al. (2015), Michinov, Morice, & Ferrières (2015), and Belmonte, Borges, & Garcia (2022) investigated different modified versions of PI and compared them against other active methodologies. In all these studies, the students' perceived satisfaction with PI was equal to or higher than for the other methodologies. In the work of Belmont, Borges, & Garcia (2022), the students stressed the relevance of the methodology for promoting group work in an online environment.

Analysis of performance in answering conceptual questions was another popular way to evaluate the impact of PI. This was achieved by observing shifts in the students' responses from incorrect to correct alternatives, after the peer discussion step in the PI sessions. The significance of these changes was typically evaluated using descriptive statistical analysis and the fractional gain (Miller et al., 2014). The same methods were used to explore data obtained for performance in final tests or questionnaires, to compare pre- and post-tests, and to evaluate the confidence of the students in their answers during PI sessions (see Table 4). The questionnaires used in the reviewed studies were from validated conceptual tests, validated questionnaires, and surveys authored by the researchers. These methods were identified by Olpak & Yilmaz (2022), where quantitative research methods, using statistical data, accounted for 68.97% of the articles reviewed.

TABLE 4. Monitored impact, data collection, methods of analysis, and results of PI implementation in the studies covered in this review.

Reference	Monitored impact	Data collection	Method of analysis	Outcomes
Bruk & Towns (2009)	Performance in conceptual questions.	Conceptual questions.	Percentual analysis	Performance increased.
Brooks & Koretsky (2011)	Performance in conceptual questions; Conceptual improvement; Students' confidence in answers.	Conceptual questions with confidence tier; Coded quality of short written responses.	Nonparametric sign test; ANOVA.	Tendency to move to the consensually popular answer, whether it was correct or not; Quality of written responses improved after peer discussion; Choosing answers before peer discussion improved confidence after peer discussion.
Michinov, Morice, & Ferrières (2015)	Performance in conceptual questions; Perception of students.	Conceptual questions; Questionnaire.	Normalized gain; ANCOVA; Normalized change.	Performance increased; Higher perceived satisfaction.
Morice et al. (2015)	Performance in conceptual questions; Perception of students.	Conceptual questions; Open question test / questionnaire.	ANCOVA; Student's t-test.	No significant changes in performance; Positive comments.
Gok & Gok (2016)	Performance in final tests; Pre- and post-tests; Perception of students.	Four tests (CAT / LSS / SEQ / textbook quantitative problems); Questionnaire.	Descriptive statistics were calculated - means and standard deviations; Fractional gains (g) (Hake, 1998); ANOVA.	Performance in the 4 questionnaires increased; Positive perception on affective and cognitive aspects of PI.
Moraes, Carvalho, & Neves (2016)	Performance in conceptual questions; Perception of students.	Questionnaire.	Participant observation.	Performance increased; Mostly positive feedback.

Aricò & Lancaster (2018)	Perception of students.	Open answer test / questionnaire.	Not informed	Mostly positive comments.
Pearson (2019)	Performance in conceptual questions; Preference (or not) for live showing during peer discussion; Perception of students.	Conceptual questions; Open answer test / questionnaire.	Percentual analysis.	Performance increased; Not showing preferred; Positive comments.
Yildirim & Canpolat (2019)	Performance in pre-tests and post-tests; Conceptual improvement; Perception of students.	Questionnaire; Test; Semi-structured interviews; Observation notes.	Statistical analysis; Content analysis.	Performance increased; Conceptual ideas were more accurate; No significant changes in perception.
Belmonte, Borges, & Garcia (2022)	Performance in conceptual questions; Performance in final test; Perception of students.	Questionnaire; Post-test.	Nonparametric statistical tests.	Performance increased; No significant changes in perception.

Bruck and Towns (2009) compared the performance of students in conceptual questions, with and without participation in peer discussions. Except for one case, all the students performed better when they collaborated with a peer. It was also observed that there was no significant difference in performance based on the type of question, classified according to the Bloom's taxonomy, Robinson/Nurrenbern, and Bretz/Smith/Nakhleh frameworks. It is well-established that in PI, there is convergence towards the correct answer after the peer discussion step (Crouch & Mazur, 2001).

Morice et al. (2015) compared the performance of two groups of students, with both participating in online (8 hours) and presential (2 hours) classes. Subsequently, one group had a single PI session (2 hours). Pre- and post-tests were applied, and the performance of the PI group was compared to that of the control group (where the students did not discuss their answers with peers). It was found that there was no significant difference between those who could and could not interact with their peers. The authors mentioned that free access to the internet could have limited peer interaction, with the blended learning style allowing the control group to be active during their learning process.

Brooks & Koretsky (2011) and Pearson (2019) examined the impact of displaying class responses during peer discussions. The former study found that showing intermediate results made no difference in terms of performance. However, seeing the responses of peers affected the self-reported confidence of the students. Furthermore, although the students tended to select the consensual answer, whether it was correct or not, those who answered correctly before the peer discussion showed improved quality of their short explanations after the discussion. It is important to note that other types of active methodologies were employed during the semester, although, they were not included in the analysis. The study by Pearson (2019) also found that showing responses during discussions increased performance, mainly due to intra-team peer discussion and discussion based on response data available from other teams. Besides the better results, it was also found that the students were strongly influenced by a polarized focus on one answer, which frequently occurred with true-false questions. Nonetheless, this encouraged further discussion about the questions. In this model of PI, only one test session was necessary. Pearson (2019) also found that the longer the time available to answer a question, the greater the number of students changing to the correct answer, even if it was not initially the most popular choice. However, this finding contradicted the results of Miller et al. (2014), which showed that correct answers were typically given before incorrect ones. The latter study also suggested that the time allowed for answers should be limited, to avoid random guessing after a certain point, since it was observed that the students tended to guess randomly after 80% of the class had responded.

Michinov, Morice, & Ferrières (2015) used normalized gain as a metric for student improvement, with ANCOVA performed to compare the gains among groups that received individual instruction, classical PI, or stepladder PI. In this case, stepladder PI resulted in the highest learning gains, followed by classical PI and individual instruction. Morice et al. (2015) used the same strategy to measure and analyze learning gain. No differences were found between the experimental and control groups, although the students in the experimental group gave high ratings to the method of instruction they participated in, as measured by a Likert scale.

Gok & Gok (2016) used four different questionnaires to collect and analyze data. The Chemistry Achievement Test (CAT), a survey created by them, covering fundamental chemistry concepts, and the Learning Strategies Survey (LSS), concerning cognitive and metacognitive skills, were implemented as pre- and post-tests for control and experimental groups. The Student Evaluation Questionnaire (SEQ), created by the authors to assess the perception of the students regarding PI, and an exam composed of selected quantitative textbook questions were implemented as final questionnaires. Descriptive statistics was used to analyze the responses of the students in the first test (CAT), with the averages showing the fractional gain (Hake, 1998). ANOVA was performed to identify statistically significant differences in post-tests, showing greater improvement of the PI group students, compared to the control group. ANOVA was also performed to evaluate differences between the answers of the PI and control groups to the LSS questionnaire. For all the components of this test, considering aspects such as organization, elaboration, critical thinking, self-regulation, and peer learning, among others, the PI group showed higher performance. For problem-solving skills, the PI students achieved higher percentages (around 90%), compared to the control group students. Additionally, most of the responses of the students to the SEQ questionnaire were positive, considering both affective and cognitive aspects. In the study of Belmonte, Borges, & Garcia (2022), the final questionnaire had four multiple-choice questions and a confidence tier (1-3). The results showed that there was no difference in concept acquisition between students who participated in PI activities and time-based learning activities.

Conceptual improvement, which is a less popular but nonetheless important approach for evaluating the implementation of PI, was applied in two studies (Brooks & Koretsky, 2011; Yildirim & Canpolat, 2019). Conceptual improvement can be defined as a measure of conceptual change based on analysis of students' written material or interviews with students. Both studies coded the students' responses, employing different tools to evaluate them. Brooks & Koretsky (2011) used statistical analysis, while Yildirim & Canpolat (2019) adopted a qualitative approach using content analysis. The former authors coded the students' explanations for their answers to multiple-choice questions. The results showed that peer discussion did not increase the correct answers, compared to students who did not have peer discussion, although those who answered correctly before the class discussion had a higher value code assigned to their short explanations after the discussion. Lower value codes were assigned to students who moved from correct to incorrect answers. The results of these studies were consistent with the findings for other disciplines (Müller et al., 2017; Olpak & Yilmaz, 2022; Vickrey et al., 2015; Young & Talanquer, 2013), where PI generally led to learning improvement and greater engagement within classes.

Considerations and implications

This systematic review provides insights into the implementation and outcomes of PI in chemistry classes in higher and basic education. The first article was published in 2009 and only in 2015, 2016, and 2019 were two articles per year identified on this topic. The findings highlight that there is much more to be explored about PI in these contexts, since there were only one or two studies in each chemistry subarea, with only two studies concerning high school chemistry classes. Nonetheless, it is notable that the methodology has been widely disseminated in STEM education. According to Olpak & Yilmaz (2022), PI is one of the most popular research-based instructional practices. Additionally, most of the studies analyzed a small sample size of students and focused on quantitative analysis. Hence, although this review only identified 10 articles concerning the use of PI in chemistry teaching, there is ample scope for new research in this area, given the observed effectiveness of PI in the teaching and learning of fundamental chemistry concepts.

Regarding the implementation of PI in classes, pre-class activities are not as popular as they should be. Since PI is associated with short times available for lectures on course topics, it is desirable for students to have prior knowledge of the class content. The lack of this step can make the learning process harder in PI sessions and generate negative evaluations of the students. In addition, changes have been made to the original PI in terms of data collection by the researchers (writing short answers to be studied later). Other changes have been more substantial, such as in stepladder PI, where the strategy adopted was not to give the students time for peer discussion between questions, but to discuss all the questions at once at the end of a sequence. Ensuring the participation of each student in peer discussion is an important variable to consider for improved learning of students, but it can be difficult to apply the stepladder PI method in most teaching contexts.

An approach that is still needed is to evaluate PI against other active methodologies in chemistry classes, since this was only performed in two studies, in different ways. This is a common practice in physics education, as exemplified by Müller et al. (2017), who described the association of JiTT with PI. Additionally, although some studies have provided more details about the management of steps of PI during its application, such as the composition of peer discussion groups, there has been little attention given to the profile or background of the students.

The outcomes of PI in chemistry classes have mainly been monitored by means of the perceptions of the students, with performance in answering conceptual questions frequently being used as an assessment tool. It is interesting to note that few studies have used validated questionnaires from the literature to assess the progress of students. Besides these two major approaches to analysis, other monitored effects were varied. However, most of the results were consistent with PI outcomes found in other science disciplines (Olpak & Yilmaz, 2022).

Another important aspect of research on teaching methodologies is the use of theoretical references. The results of empirical studies would be better supported if theoretical references were used in designing and discussing the work. In this review, however, only a few papers cited a broad theoretical reference, such as "socio-constructivist theory", to support the findings, and there was little use of such references in discussion of the outcomes (Nakhleh, 1992). This does not disqualify the findings, but such discussions would certainly contribute further to the literature and to teaching practices. This gap was also noted in the review conducted by Müller et al. (2017).

The results of the research analyzed here can be considered significant, but it is important to note that they should not be generalized to all chemistry subareas and contexts, due to the limited number of studies and small sample sizes. Therefore, more research is needed to investigate the outcomes of PI in diverse and broad chemistry class contexts. The present review, in answering the research questions, gives a comprehensive assessment of the potential of PI applied in chemistry education. The information provided here can support chemistry teachers, instructors, and professors in their decisions to use Peer Instruction (and its variations), promoting evidence-based implementation of PI in chemistry classes worldwide.

References

Aricò, F. R. & Lancaster, S. J. (2018). Facilitating active learning and enhancing student self-assessment skills. *International Review of Economics Education*, *29*, 6-13. https://doi.org/10.1016/j.iree.2018.06.002

Belmonte, I. S., Borges, A. V., & Garcia, I. T. S. (2022). Adaptation of physical chemistry course in COVID-19 period: reflections on peer instruction and team-based learning. *Journal of Chemical Education*, *99*(6), 2252-2258. http://dx.doi.org/10.1021/acs.jchemed.1c00529

Bernardi, F. M. & Pazinato, M. S. (2022). The case study method in chemistry teaching: A systematic review. *Journal of Chemical Education*, *99*(3), 1211-1219. http://dx.doi.org/10.1021/acs.jchemed.1c00733

Børte, K., Nesje, K., & Lillejord, S. (2023). Barriers to student active learning in higher education. *Teaching in Higher Education*, *28*(3), 597-615. https://doi.org/10.1080/13562517.2020.1839746

Brooks, B. J. & Koretsky, M. D. (2011). The influence of group discussion on students' responses and confidence during peer instruction. *Journal of Chemical Education*, *88*(11), 1477-1484. http://dx.doi.org/10.1021/ed101066x

Bruck, A. D. & Towns, M. H. (2009). Analysis of classroom response system questions via four lenses in a general chemistry course. *Chemistry Education Research and Practice*, *10*(4), 291-295. http://dx.doi.org/10.1039/B920834H

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

Crouch, C. H., Watkins, J., Fagen, A. P., & Mazur, E. (2007). Peer instruction: Engaging students one-on-one, all at once. *Research-based Reform of University Physics*, *1*(1), 40-95. http://dx.doi.org/10.1119/RevPERv1.1.3

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410-8415. https://doi.org/10.1073/pnas.1319030111

Gok, T. & Gok, O. (2016). Peer instruction in chemistry education: Assessment of students' learning strategies, conceptual learning and problem solving. In *Asia-Pacific Forum on Science Learning and Teaching* (pp. 1-21). Hong Kong, CN.

Graffam, B. (2007). Active learning in medical education: strategies for beginning implementation. *Medical Teacher*, *29*(1), 38-42. http://dx.doi.org/10.1080/01421590601176398

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64-74. http://dx.doi.org/10.1119/1.18809

Heiner, C. E., Banet, A. I., & Wieman, C. (2014). Preparing students for class: How to get 80% of students reading the textbook before class. *American Journal of Physics*, *82*(10), 989-996. http://dx.doi.org/10.1119/1.4895008

Lucas, A. R. (2012). Using WeBWorK, a web-based homework delivery and grading system, to help prepare students for active learning. *Primus*, *22*(2), 97-107. https://doi.org/10.1080/10511970.2010.497834

Mazur, E. (1997). Peer instruction: Getting students to think in class. In *AIP Conference Proceedings* (Vol. 399, No. 1, pp. 981-988). Cambridge, MA.

Mazur, E. (1999). *Peer Instruction: A User's Manual*. Upper Saddle River, NJ: Prentice Hall.

Michinov, N., Morice, J., & Ferrières, V. (2015). A step further in peer instruction: Using the stepladder technique to improve learning. *Computers & Education*, *91*, 1-13. http://dx.doi.org/10.1016/j.compedu.2015.09.007

Miller, K., Lasry, N., Lukoff, B., Schell, J., & Mazur, E. (2014). Conceptual question response times in peer instruction classrooms. *Physical Review Special Topics - Physics Education Research*, *10*(2), 020113. http://dx.doi.org/10.1103/PhysRevSTPER.10.020113

Moraes, L. D. de M., Carvalho, R. S., & Neves, Á. J. M. (2016). O Peer Instruction como proposta de metodologia ativa no ensino de química. *The Journal of Engineering and Exact Sciences*, *2*(3), 107-131. http://dx.doi.org/10.18540/jcecvl2iss3pp107-131

Morice, J., Michinov, N., Delaval, M., Sideridou, A., & Ferrières, V. (2015). Comparing the effectiveness of peer instruction to individual learning during a chromatography course. *Journal of Computer Assisted Learning*, *31*(6), 722-733. https://doi.org/10.1111/jcal.12116

Müller, M. G., Araujo, I. S., Veit, E. A., & Schell, J. (2017). Uma revisão da literatura acerca da implementação da metodologia interativa de ensino Peer Instruction (1991 a 2015). *Revista Brasileira de Ensino de Física*, *39*. http://dx.doi.org/10.1590/1806-9126-rbef-2017-0012

Nakhleh, M. B. (1992). Why some students don't learn chemistry: Chemical misconceptions. *Journal of Chemical Education*, *69*(3), 191. https://doi.org/10.1021/ED069P191

Novak, G., Gavrin, A., Christian, W., & Patterson, E. (1999). *Just-in-Time Teaching: Blending Active Learning with Web Technology.* Upper Saddle River, NJ: Prentice Hall.

Olpak, Y. Z. & Yilmaz, R. (2022). Review of Trends in Peer Instruction: Bibliometric Mapping Analysis and Systematic Review. *Journal of Learning and Teaching in Digital Age*, *7*(1), 42-50. https://doi.org/10.53850/joltida.978824

Pearson, R. J. (2019). Exploring Peer Instruction: Should Cohort Clicker Responses Appear During or After Polling? *Journal of Chemical Education*, *96*(5), 873-879. http://dx.doi.org/10.1021/acs.jchemed.9b00035

Pollozi, S., Haddad, I., Tyagi, A., Mills, P., & McGregor, D. (2019). Using clicker-based group work facilitated by a modified peer instruction process in a highly successful flipped general chemistry classroom. In *Active Learning in General Chemistry: Whole-class Solutions* (pp. 35-53). American Chemical Society.

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, *93*(3), 223-231. http://dx.doi.org/10.1002/j.2168-9830.2004.tb00809.x

Sadler, I. (2012). The challenges for new academics in adopting student-centred approaches to teaching. *Studies in Higher Education*, *37*(6), 731-745. http://dx.doi.org/10.1080/03075079.2010.543968

Simpson, R. D. & Oliver, J. S. (1990). A summary of major influences on attitude toward and achievement in science among adolescent students. *Science Education*, *74*(1), 1-18. https://psycnet.apa.org/doi/10.1002/sce.3730740102

Vickrey, T., Rosploch, K., Rahmanian, R., Pilarz, M., & Stains, M. (2015). Research-based implementation of peer instruction: A literature review. *CBE—Life Sciences Education*, *14*(1), es3. http://dx.doi.org/10.1187/cbe.14-11-0198

Yıldırım, T. & Canpolat, N. (2019). An investigation of the effectiveness of the peer instruction method on teaching about solutions at the high-school level. *Egitim ve Bilim*, *44*(199). http://dx.doi.org/10.15390/EB.2019.7966

Young, K. K. & Talanquer, V. (2013). Effect of different types of small-group activities on students' conversations. *Journal of Chemical Education*, *90*(9), 1123-1129. http://dx.doi.org/10.1021/ed400049a

Zhang, P., Ding, L., & Mazur, E. (2017). Peer Instruction in introductory physics: A method to bring about positive changes in students' attitudes and beliefs. *Physical Review Physics Education Research*, *13*(1), 010104. http://dx.doi.org/10.1103/PhysRevPhysEducRes.113.010104

Recepción: 19/07/2024 ▪ Aceptación: 21/09/2024.