



Cite this: *Chem. Educ. Res. Pract.*, 2020, 21, 1083

## Exploring the impact of the reasoning flow scaffold (RFS) on students' scientific argumentation: based on the structure of observed learning outcomes (SOLO) taxonomy

Xiuling Luo,<sup>a</sup> Bing Wei, <sup>\*b</sup> Min Shi<sup>a</sup> and Xin Xiao<sup>a</sup>

Using the Structure of Observed Learning Outcomes (SOLO) taxonomy as the analytic framework, this study examined the impact of the reasoning flow scaffold (RFS) on students' written arguments. Two classes with a total of 88 10th grade students in a school participated in this study. One class, set as the experimental group, was taught scientific argumentation with RFS whereas the control class received conventional argumentation teaching. They all experienced three argument assignments of writing scientific arguments and the measurement task before and after the teaching intervention. The results of data analysis showed that after teaching intervention, students in the experimental group performed significantly better than those in the control group on evidence and rebuttal while there were no significant differences on claim or reason between the two groups. Some implications and suggestions are provided in the last part of this paper.

Received 22nd November 2019,  
Accepted 19th May 2020

DOI: 10.1039/c9rp00269c

rsc.li/cerp

### Introduction

Over the past decade, argumentation has received considerable attention in curriculum initiatives and reforms of science education across the world. The European Union (EU) officially recommended that scientific argumentation should be incorporated as a set of key competencies for lifelong learning (EU, 2006). Later on, 'engaging in argument from evidence' was suggested by The National Research Council (NRC, 2012) as one of the scientific and engineering practices. The Program for International Student Assessment (PISA) considers the ability to use scientific evidence to support claims, which is a fundamental skill for engaging in scientific argumentation, to be a key competency (Organization for Economic Cooperation and Development [OECD], 2006, 2013). Scientific argumentation has also gained attention in China in recent years. For instance, in the latest chemistry curriculum reform at the stage of senior high school in China, evidence-based reasoning has been suggested as one of the five dimensions of the chemistry core competency in the view of argumentation (Wei, 2019, 2020).

The notion of argumentation refers to the process of creating arguments and is interpreted as a scientific practice in which scientific claims are justified or estimated based on empirical or theoretical evidence (Jiménez-Aleixandre and Erduran, 2007). In recent decades, student argumentation has been studied across a range of ages in two basic contexts, oral and written

argumentation (Erduran *et al.*, 2004). For example, von Aufschnaiter *et al.* (2008) investigated junior high school students' processes of argumentation and cognitive development in socio-scientific lessons. Sampson, *et al.* (2011) examined how a series of laboratory activities influenced the quality of scientific arguments provided by 10th-grade students. In particular, many studies have revealed that students have difficulties in using evidence to justify claims and rebutting counter arguments (Osborne *et al.*, 2004; Berland and Reiser, 2009; McNeill, 2011). For school students, as Osborne *et al.* (2004) observed, constructing a good argument is not a simple task, and they need sufficient support and clear guidance to help them build their sense of what an effective argument is.

As a teaching strategy, scaffolding has been used to help students construct arguments, mainly in the form of charts or diagrams. For example, Nussbaum (2002) designed an argumentation scaffold in the form of a diagram with a box representing argument elements and arrows indicating the relationship of argument elements. Chin *et al.* (2010) designed a guided-TAPPING pattern related to the issue of genetically modified food, helping sixth-grade students construct arguments. They found that the guided-TAPPING pattern could help students organize their thinking and improved their writing in the argument construction. Later, Chin *et al.* (2016) used their guided-TAPPING pattern in helping students learn about the global climate change issue, and found that there were significant improvements in students' argumentation and science understanding. However, few studies have been conducted to specifically examine the effectiveness of this teaching strategy on improving the quality of students' scientific argumentation. In this study, the reasoning flow scaffold (RFS) was developed on the

<sup>a</sup> School of Chemistry, South China Normal University, Guangzhou, Guangdong, China

<sup>b</sup> Faculty of Education, University of Macau, Taipa, Macau, China.

E-mail: bingwei@um.edu.mo

basis of Toulmin's (1958) argument pattern with the purpose of helping students make arguments more clearly and logically. The present study was purported to provide empirical data to answer the two research questions: To what extent does the RFS teaching strategy promote students' written arguments? How might the SOLO taxonomy serve as an analytical framework to evaluate student-generated argumentation?

## Literature review

### Scaffolding

The metaphor of scaffolding has been used to reflect the way adult support is adjusted as the child learns and is ultimately removed when the learner can stand alone (Wood *et al.*, 1976). While illuminating the nature of adult interactions in child learning, these researchers identified six key elements of scaffolding, they are: (1) recruitment (enlisting the learners interest and adherence to the requirements of the task); (2) reduction in degrees of freedom (simplifying the task so that feedback is regulated to a level that could be used for correction); (3) direction maintenance (keeping the learner in pursuit of a particular objective); (4) marking critical features (accentuating some and interpreting discrepancies); (5) frustration control (responding to the learners emotional state); and (6) demonstration, or modelling a solution to a task (p. 98). Saye and Brush (2002) divided scaffolds into two types – soft and hard. Soft scaffolds refer to just-in-time support provided by teachers or peers to each student dynamically; but it is difficult for them to meet the needs of the whole class immediately. Hard scaffolds which can make up for this defect are provided by computer software or paper-based cognitive tools. Hard scaffolds are static supports that can be anticipated and planned based on typical learning difficulties with a task ahead of teaching (Saye and Brush 2002; Belland *et al.*, 2008). In the field of science education, scaffolding has proved to be effective to help students integrate more evidence with scientific argumentation (Oh and Jonassen, 2007; Iordanou and Constantinou, 2015) and understand the criteria of good arguments (Sandoval, 2003; Erduran *et al.*, 2004; McNeill *et al.*, 2006; von Aufschnaiter *et al.*, 2008).

Guided by the seminal work of Wood and colleagues (1976) and with the support of previous research (Nussbaum, 2002; Chin *et al.*, 2010; Chin *et al.*, 2016), RFS in this study focused on the reasoning process of argumentation, appearing as a box-and-arrow diagram with boxes representing argument components and arrows representing the logical relationships in the argumentation (see Appendix 1). These explicit architectural features tend to help students familiarize with the structure of arguments and generate good arguments. According to Saye and Brush (2002)'s classification, it is one kind of hard scaffold based on anticipated student needs. That is, to respond to students' possible argumentation difficulties, the basic structure of RFS has been designed in advance and the forms of RFS are varied across specific assignments.

### The SOLO taxonomy

To evaluate the quality of student arguments, the Structure of Observed Learning Outcomes (SOLO) taxonomy proposed by Biggs and Collis (1982) was adopted in this study. As a systematic way of

classifying and describing the range of students' performances, the SOLO taxonomy has two major features: modes of thinking and levels of response (complexity within each mode) (Biggs and Collis, 1982). According to the SOLO taxonomy, understanding is considered an increase in the number and complexity of connections students construct as they progress from novice to expert. It comprises five levels of increasing sophistication, namely: Pre-structural, Uni-structural, Multi-structural, Relational and Extended Abstract, with each level covering and transcending the previous one (Caniglia and Meadows, 2018). The meanings of the five levels are explained as follows:

(1) Pre-structural (P). Students are frequently distracted or misled by irrelevant elements of the situation. The responses are completely inconsistent with the question, or the question is reworded.

(2) Uni-structural (U). Students are focusing on one aspect of the task, using only one piece of information, fact, or idea, obtained directly from the problem. They may have limited knowledge on the topic or do not see connections between ideas, and provide a fact or concept in isolation.

(3) Multi-structural (M). Students may know a few facts about the topic but each piece of information is used separately with no integration of the ideas. And two or more aspects of the task are viewed discretely and treated separately.

(4) Relational (R). Moving toward a higher level of thinking, students are able to provide explanations that link relevant details. At least two separate pieces of information, facts, or ideas, work together to explain several ideas pertaining to a topic, and the components are integrated into a coherent whole structure with consistency.

(5) Extended abstract (E). In the most complex stage, students are able to derive a general principle from the integrated data through reproduction and evaluation, and they can apply it to new situations. The response goes beyond the given information and deduces a more general rule or proof that applies to other scenarios.

The SOLO taxonomy has established a qualitative evaluation system based on hierarchical description, which is an effective diagnostic tool for students' learning process (Minogue and Jones, 2009). It describes levels of progressively complex understanding over five general stages that are intended to be relevant to all subjects within all disciplines and has been used widely in many studies (*e.g.*, Olive, 1991; Potter and Kustra, 2012; Caniglia and Meadows, 2018).

### Analytic framework of scientific argumentation

Toulmin's (1958) model of argumentation is composed of six interactive components: claim, data, warrant, backing, qualifier, and rebuttal. This model has been widely used to both introduce and assess argumentation in science education (Bell and Linn, 2000; Erduran *et al.*, 2004; Osborne *et al.*, 2004; McNeill *et al.*, 2006; Sadler and Fowler, 2006). In the process of application, this model has been revised by researchers in two aspects: reduction of the number of components evaluated in arguments and inclusion of measurement of scientific content (Cooper and Oliver-Hoyo, 2016). According to Cooper and Oliver-Hoyo (2016), condensing the number of components evaluated in arguments has two advantages. First, the language used in the prompts for argument construction

can be broken into sections to simplify the process for students. Second, lowering the number of components simplifies the evaluation of the constructed arguments. In their study, for instance, McNeill *et al.* (2006) condensed the six components to three: claim, evidence, and reasoning, which are frequently cited as essential elements in the evaluation of scientific argumentation (Sandoval, 2003; Ruiz-Primo *et al.*, 2010; Braaten and Windschitl, 2011; Sampson, *et al.*, 2011). However, we concur with Osborne *et al.*, (2004)'s contention that rebuttal is also an essential component in argument evaluation for better quality. This is the reason that we adopted four key components, *i.e.* claims, evidence, reasons and rebuttals, in this study in assessing students' arguments.

Furthermore, Toulmin's model has proved to be difficult to use to verify the correctness of arguments. For example, although a student argument can be considered relatively strong according to Toulmin's model, the content may be inaccurate from a scientific perspective (Sampson and Clark, 2008). To assess science content in arguments, for instance, Sandoval's (2003) proposed guidance for assessing the conceptual quality of students' arguments and the sufficiency of the evidence cited by students. However, the coding system of Sandoval's framework was embedded in the science content of natural selection, a biological topic. For application to other content areas, as suggested by Sampson and Clark (2008), a domain-general analytic framework is needed. This is the reason that the SOLO taxonomy is adopted in the present study.

According to Sampson and Clark (2008), to assess the quality of scientific arguments, three critical issues should be addressed, they are: (1) the structure or complexity of the argument (*i.e.*, the components of an argument); (2) the content of an argument (*i.e.*, the accuracy or adequacy of the various components in the argument when evaluated from a scientific perspective); and (3) the nature of the justification (*i.e.*, how ideas or claims are supported or validated within an argument). The SOLO taxonomy which characterizes levels of complexity in student reasoning according to the number of components of arguments and degree

of integration, has been used to develop theoretically grounded scales for measuring scientific reasoning (Claesgens, *et al.*, 2009; Brown, *et al.*, 2010a; 2010b; Bernholt and Parchmann, 2011). Based on the combination of the SOLO taxonomy and the four components of argumentation, we have established a rubric for assessing student-generated scientific arguments (see Table 1).

## Methods

### Research design

A quasi-experiment was conducted to examine the impact of RFS teaching on students' written arguments. Two 10th grade classes in a high school were selected as control group and experimental group respectively. Students in the experimental group were taught with RFS teaching strategy while students in the control group received conventional argumentation teaching that did not use the RFS teaching strategy. Different teaching strategies for teaching scientific argumentation were the independent variables, and the dependent variable was the students' levels in constructing written arguments of the measurement task in pre- and post-tests. This research project went through a rigorous ethics review at the first author's university before the data were collected. The participating chemistry teacher had signed a consent letter for a number of research activities during the period of this study.

### Participants

Students from the two selected classes took part in this study on a voluntary basis. They were between the ages of 15 and 17. The students were informed of the general research objective of this study but they were not aware of whether they were in the control or experimental group. The students were told that their answers would have no influence on their final course grades, and that they could drop out at any time. Students who completed all of the three written argument assignments and the measurement task before and after the teaching intervention

**Table 1** The rubric of assessing students' written arguments based on the SOLO taxonomy

Level	Component			
	Claim	Evidence	Reason	Rebuttal
P (level 1)	Irrelevant, incorrect or illogical claims; reworded question; or no claim.	Inaccurate, irrelevant or illogical evidence; or no evidence.	Inaccurate, irrelevant or illogical reason; or no reason.	Inaccurate, irrelevant or illogical rebuttal; or no rebuttal.
U (level 2)	Provide a correct claim.	Provide a correct, relevant and logical piece of evidence.	Provide a correct, relevant and logical reason.	Provide a relevant, correct and logical rebuttal.
M (level 3)	Present more than two correct and logical claims, but they are separated.	Present more than two correct and logical pieces of evidence to support claims, but there is no connection among evidence.	Present more than two correct and logical reasons to show how the evidence can support the claim, but there is no connection among reasons.	Present more than two rebuttals, but there is no connection among these rebuttals.
R (level 4)	Present more than two correct, logical claims which work together to answer the question, or the relation among these claims is explained.	Present more than two correct and relevant pieces of evidence, and they are integrated into a coherent whole with consistency.	Present more than two correct, logical and relevant reasons, and multiple reasons are integrated, or the relation among them is explained.	Present more than two correct, logical and relevant rebuttals, and they work together, or the relation among them is explained.
E (level 5)	Offer valuable, integrated and opening claims at a higher level of abstraction, which would be suitable for new situations.	Present evidence from multiple perspectives, summarizing the links among them, and refer to new evidence that the data or chart does not show.	Present more than two correct, logical and relevant reasons, and show new reasons related to upper level concepts or theories.	Present more than two correct, logical and relevant rebuttals, and they work together, which can be applied to new situations.

were considered as subjects in this study. The total number of subjects in this study was 88, including 45 students in the experimental group (24 males, 21 females) and 43 students in the control group (22 males, 21 females).

### Procedure

The teaching intervention was provided by one of the authors, who is familiar with scientific argumentation and RFS teaching. To rule out the possible negative effects that might be caused by the change of the instructor, both groups were taught by the new instructor for two weeks ahead of the teaching intervention to let students adapt to her teaching style. The two groups were arranged in the same class schedules, that is, pre-tests, post-tests and three written argument assignments were carried out at almost the same time. The study lasted for about sixteen weeks. The outline of this study is demonstrated in Table 2.

In the third week, the two groups of students took the pre-test in class. In the test, they were asked to complete the measurement task in 40 minutes individually. Their copies were collected on time. Since the same task was used in the post-test, no guidance or feedback was provided to the students after the pre-test.

To improve the quality of participants' writing arguments as much as possible, each assignment was carried out after the related chemistry content had been taught. The teaching intervention materials used in both groups were designed by the authors of this article. The experimental groups of students were introduced to RFS and used it to complete three assignments of writing arguments on three different topics: (1) change in weight of sodium hydroxide in air, (2) cations determination in gluconate complex preparations, and (3) color change in the fountain of chlorine and sodium hydroxide solution. The control groups of students completed the same assignments but they were not introduced to RFS.

For each assignment, the worksheet provided to the students included two parts. Part 1 was information on the topic of the argument, including experimental design, phenomena and data, which was the same for the two groups. In Part 2, students were required to write arguments based on Part 1. Part 2 was presented to the two groups of students in different ways. The experimental group of students were asked to establish arguments by completing the RFS box while the control group of students were asked to complete the argument components (evidence, claim, reason, and rebuttal) without the RFS box. In each RFS, there were mainly five connected components: evidence, reasons, rebuttals, preliminary claims, and final claims. The form of each RFS was different, that is to say, the

connection between boxes and arrows was varied dependent on the topic of the arguments. Before formal teaching, each RFS was tested by six 10th grade students, and was reviewed by two chemistry educators and a school chemistry teacher. An example of written argument assignment is given in Appendix 1, the topic of which was determining cations in gluconate complex preparations, and an example of written argument assignment by an experimental student is provided in Appendix 2.

The three completed assignments were collected and the teacher provided explicit and detailed feedback. Then, the students modified and improved their written arguments based on the teacher's comments. Finally, standard answers were provided for the two groups of students: standard answers in the RFS structured form for the experimental group and standard answers without the RFS for the control group. It should be noted that both groups of students completed and modified their three assignments in the form of homework. Three weeks after the teaching intervention, the post-test was administered to both groups of students using the same measurement task in class. Students were required to answer the questions individually within 40 minutes. After they had finished the test, all the copies were collected on time.

### Measurement instrument

The measurement instrument was an argument construction task about the corrosion rate of iron nails and the influencing factors, which was adapted from that used in a doctoral dissertation (Han, 2016) with references to other materials (Wright and Stephen, 2005; Moraes, *et al.*, 2015; Sanders, *et al.*, 2018). Corrosion of iron is a common phenomenon in daily life and it is also a very complex electrochemical process in which iron, oxygen and water react together to form rust. The corrosion of iron is represented in many different ways in the chemistry course. It is a typical example that is used to understand the law of conservation of mass or redox chemistry (Wong *et al.*, 2012). According to Cornell and Schwertma (2003), the possible reaction mechanism of the corrosion of iron is depicted in Fig. 1.

The content validity of the test was established through the expert panel methodology (Aydeniz *et al.*, 2012). Three experts participated in the construction of the questions. All of them hold a chemistry degree, and some of them had advanced degrees (PhD) in science education or chemistry. They all have previously conducted

Table 2 Outline of the study

Week	Activity	Time (min)
3rd	Pre-test (in class)	40
5th	Argumentation teaching	40
6th	Assignment-1 (homework)	
7th	Feedback-1 (classroom activity)	20
9th	Assignment-2 (homework)	
10th	Feedback-2 (classroom activity)	20
12th	Assignment-3 (homework)	
13th	Feedback-3 (classroom activity)	20
16th	Post-test (in class)	40

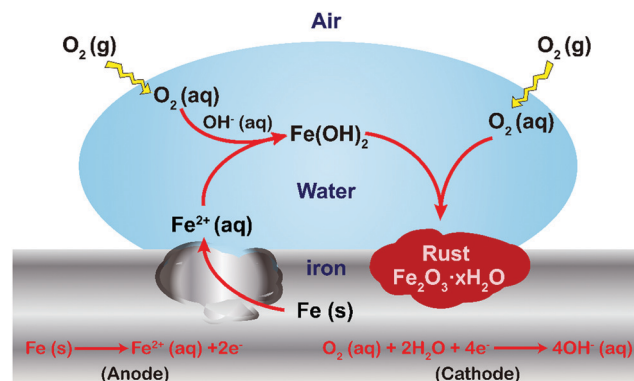


Fig. 1 Reaction mechanism of corrosion of iron.

research in chemical education, and have been teaching in university or high school for more than ten years. To determine the difficulties and time required to complete the task, ten students with similar educational background with the participants of this study were invited to complete the task. Based on their feedback, the measurement task was revised to make it more easily understood. The authors developed and evaluated the experiments and questions in the measurement task iteratively until a consensus was established on the correctness of the content, the appropriate difficulty level, and the ease of language used in the questions. The measurement task is shown in Appendix 3.

### Data analysis

Each student's answers to the questions in the measurement task of pre- and post-tests were coded based on the rubric shown in Table 1. They were first classified into the four components of argumentation (evidence, claim, reason, and rebuttal), then assigned to one level of the SOLO taxonomy for each component: level 1 (Pre-structural), level 2 (Uni-structural), level 3 (Multi-structural), level 4 (Relational), and level 5 (Extended abstract). In order to explain the coding process, we provide answers of a student from the experimental group in pre- and post-tests as follows with a focus on the component of 'evidence':

Pre-test: *The iron nail in No. 2 test tube was corroded fastest. The iron nail in No. 5 was corroded most slowly.* (Level 1, Pre-structural)

Post-test: *A small amount of rust appeared in No. 1 and No. 5 test tubes, but much less than that in No. 2 and No. 4 ones. In No. 2 test tube, a lot of dark black solid fell off, where the iron nail corrosion was the most serious. The surface of the iron nail in No. 3 turned reddish brown.* (Level 4, Relational)

The component of 'evidence' in the pre-test provided by the student was assigned to level 1 (Pre-structural) because he merely repeated a claim as evidence but did not distinguish between evidence and claim. While in the post-test, the student was able to describe various experimental phenomena in the five test tubes. Besides, he was able to compare those phenomena he observed. Thus, the answer was assigned to level 4 (Relational).

Two of the authors graded all students' responses independently with the purpose of reducing the subjectivity and ensuring reliability. For those inconsistencies, discussions were held among the researchers to reach a consensus. Once the 88 subjects' answers were coded with the four argument components and the five levels of SOLO taxonomy, the frequency of students at different SOLO levels for each argument component was counted. Considering the different numbers of students in the experimental group ( $n = 45$ ) and the control group ( $n = 43$ ), the percentages were determined to examine the effect of the RFS strategy in improving students' argumentation.

This study was conducted in the Chinese context. The teaching materials, measurement tasks, and students' written arguments were all presented in Chinese. In the stage of data analysis, the coding work was conducted according to the rubric we established (see Table 1). Students' answers reported in this article were originally in Chinese and translated into English. When preparing this article, the results of this study were translated from Chinese into English by the first two authors. To ensure the quality of the translation quality, an English expert, who is a Chinese native speaker, was invited to proofread all of the translations.

## Results

For each of the argument components, the frequency and percentages of students in the two groups at the five SOLO levels in the pre-test and post-test are presented in Table 3. To demonstrate the effectiveness of the teaching intervention, the percentages of the students from the two groups at each level in the pre-test and post-test for each of the four argument components are presented in Fig. 2.

Based on Fig. 2, it can be seen that most of the students in both groups were at lower levels (levels 1, 2 and 3) in the pre-test for the argument components of claim, evidence and reason. Of rebuttal, all of the two groups of students were at the three lower levels (levels 1, 2 and 3), and no student was at levels 4 and 5 in the pre-test. After argumentation teaching, the percentages of level 4 and level 5 of all the four components of the two groups of students were increased and the percentages of

**Table 3** The frequency and percentages of students in the two groups at five SOLO levels in pre- and post-tests for each argument component

		SOLO level	Level 1 (pre-structural)	Level 2 (uni-structural)	Level 3 (multi-structural)	Level 4 (relational)	Level 5 (extended abstract)
Claim	E	Pre-test	0 (0)	13 (29)	25 (56)	7 (16)	0 (0)
		Post-test	0 (0)	1 (2)	20 (44)	18 (40)	6 (13)
	C	Pre-test	1 (2)	13 (30)	23 (54)	6 (14)	0 (0)
		Post-test	0 (0)	4 (9)	22 (51)	14 (33)	3 (7)
Evidence	E	Pre-test	1 (2)	11 (24)	26 (58)	6 (13)	1 (2)
		Post-test	0 (0)	2 (4)	14 (31)	21 (47)	8 (18)
	C	Pre-test	2 (5)	9 (21)	25 (58)	7 (16)	0 (0)
		Post-test	0 (0)	4 (9)	22 (51)	14 (33)	3 (7)
Reason	E	Pre-test	12 (27)	20 (44)	11 (24)	2 (4)	0 (0)
		Post-test	1 (2)	10 (22)	19 (42)	11 (24)	4 (9)
	C	Pre-test	14 (33)	16 (37)	12 (28)	1 (2)	0 (0)
		Post-test	3 (7)	12 (28)	20 (47)	7 (16)	1 (2)
Rebuttal	E	Pre-test	26 (58)	16 (36)	3 (7)	0 (0)	0 (0)
		Post-test	11 (24)	20 (44)	9 (20)	4 (9)	1 (2)
	C	Pre-test	28 (65)	14 (33)	1 (2)	0 (0)	0 (0)
		Post-test	14 (33)	25 (58)	3 (7)	1 (2)	0 (0)

Note: the percentage (%) is in bracket. E: experimental group ( $n = 45$ ); C: control group ( $n = 43$ ).

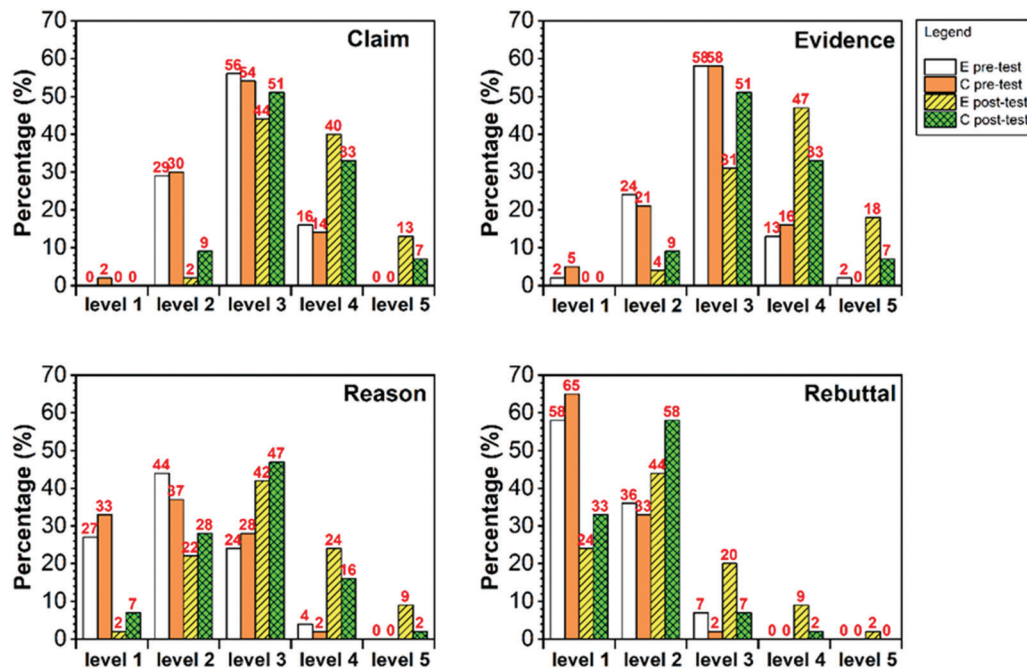


Fig. 2 The percentages of students of the two groups at five SOLO levels for claim, evidence, reason and rebuttal.

levels 1, 2 and 3 were decreased in the post-test. However, more experimental group students reached level 4 and level 5. Of the component of claim, 40% of experimental group students reached level 4 and 13% reached level 5, compared with 33% of the control group students reaching level 4 and 7% reaching level 5. As for the percentages in evidence, 47% of the experimental group students achieved level 4, compared with 33% of the control group students reaching that level. In particular, 18% of the experimental group students reached the highest level, over two times more than those of the control group students, the proportion of which was only 7%. In giving reasons, the percentage of level 4 was improved from 4% to 24% and the percentage of level 5 was from 0 to 9% in the experimental group, while in the control group of students, the percentage was from 2% to 16% (level 4) and from 0 to 2% (level 5). Of rebuttal, 9% of the experimental group got level 4 compared with 2% of the control group students reaching the same level. More significantly, except for evidence, there was no student at the top level in the pre-test, but 13% (in claim), 9% (in reason) and 2% (in rebuttal) of the experimental group of students were promoted to the

top level, compared with 7%, 2% and 0 respectively in the control group of students who had never been introduced to and used RFS.

It should be noted that regardless of all the components in the pre-test or the post-test, no more than half of the students were at level 4, which requires students to provide and integrate at least two relevant details into a coherent whole structure and maintain consistency. No more than 20% of students reached level 5, which requires students to go beyond the given information and deduce an integrated whole to generalize to new situations.

In order to investigate whether statistically significant differences existed between the two groups in pre- and post-test scores, the Mann-Whitney  $U$  test, which is a non-parametric test, was employed. The results of the Mann-Whitney  $U$  test are shown in Table 4.

As indicated in Table 4, there was no significant difference between the two groups for claim ( $U = 920.00$ ,  $p = 0.659$ ), evidence ( $U = 940.00$ ,  $p = 0.798$ ), reason ( $U = 930.00$ ,  $p = 0.740$ ) or rebuttal ( $U = 883.50$ ,  $p = 0.412$ ) prior to the teaching intervention. There were significant differences between the two groups

Table 4 Results of the Mann-Whitney  $U$ -test between the two groups

	Control group ( $n = 43$ )		Experimental group ( $n = 45$ )		$U$ value	$p$ value
	Mean rank	Sum of ranks	Mean rank	Sum of ranks		
Claim-pre	43.40	1866.00	45.56	2050.00	920.00	0.659
Evidence-pre	45.14	1941.00	43.89	1975.00	940.00	0.798
Reason-pre	43.63	1876.00	45.33	2040.00	930.00	0.740
Rebuttal-pre	42.55	1829.50	46.37	2086.50	883.50	0.412
Claim-post	40.37	1736.00	48.44	2180.00	790.00	0.106
Evidence-post	38.19	1642.00	50.53	2274.00	696.00	0.015 <sup>a</sup>
Reason-post	40.02	1721.00	48.78	2195.00	775.00	0.088
Rebuttal-post	39.45	1696.50	49.32	2219.50	750.50	0.048 <sup>a</sup>

<sup>a</sup>  $p < 0.05$ .

for evidence ( $U = 696.00$ ,  $p = 0.015$ ) and rebuttal ( $U = 750.50$ ,  $p = 0.048$ ) while there was no significant difference for claim ( $U = 790.00$ ,  $p = 0.106$ ) or reason ( $U = 775.00$ ,  $p = 0.088$ ) after the teaching intervention. That is to say, students in the experimental group have constructed higher quality evidence and rebuttal arguments than those in the control group.

## Discussion and implications

This study has examined the impact of the reasoning flow scaffold (RFS) on high school chemistry students' written arguments. The results showed that RFS instruction had significantly improved students' ability of providing evidence and rebuttal, though the improvement of claim and reason was not significant. 'Evidence' is considered as an essential component of skilled scientific argumentation (Toulmin, 1958). The significant improvement of the level of evidence in the experimental group of students could be explained by the features of scaffolding. As shown earlier, the RFS strategy dissects the complex work procedures and turns them into clear basic elements that make it easier for students to understand the argument components and enforce them to identify and use evidence. Moreover, RFS provides a streamlined logical structure, helping students think logically and think about how to write an argument with evidence.

With regard to 'rebuttal', although both groups of students were mostly at lower levels according to the SOLO taxonomy, our results have demonstrated that the quality of students' rebuttal improved significantly in the experimental group after RFS instruction. No student in the control group was able to provide a rebuttal at the highest SOLO level, while some in the experimental group reached the top level, with the proportion being 2%. As Ryu and Sandoval (2012) indicated in their study, students often ignored rebuttals or failed to make effective rebuttals. A number of studies have also drawn a similar conclusion that it is even more difficult to propose rebuttal arguments than to generate claims, warrants, and backing arguments without assistance (e.g., Clark and Sampson, 2007; Yeh and She, 2010; Katchevich, *et al.*, 2013). In our study, RFS has offered an explicit space of 'rebuttal' for students to elaborate on. Moreover, RFS divided the 'claim' into two parts, initial and final claims, which could assist students to reflect on and think carefully about the possibility of falsification. As such, the structured argumentation scaffold used in this study could encourage students to ask for a higher level of rebuttals.

As for 'claim', there was no significant difference between the two groups in the post-test. As Novak and Treagust (2018) reported, it is difficult to change a person's view or generate a new claim over a period of time. Chang and Chiu (2008) used the Lakatos' framework to evaluate students' written arguments about socio-scientific issues, and found that it was difficult to change the 'hard core' that is a student's own claim during a short period of time. Our findings were similar to those in the above studies. Thus, how to change students' claims still needs more exploration in the future. With regard to 'reason', the two groups of students both improved after teaching intervention, but there was no significant difference between them. This may be due to the fact that the chemistry concepts and theories they learned were almost the same for the two groups. As implied in

Driver *et al.* (2000)'s study, warrants (the same with reasons in the present study), referring to rules, principles, *etc.*, could be mobilized to justify the connections between the data and the knowledge claim, or conclusion. That is to say, when students possess a certain degree of theoretical knowledge, they may be able to propose some reasons.

The analysis of students' arguments can provide a great deal of information about students' understanding of scientific content, reasoning, epistemological commitments, and their ability to communicate and prove ideas to others (Sampson and Clark, 2008). In the present study, we have demonstrated that the SOLO taxonomy could be used as an analytical framework for evaluating student-generated argumentation. As evidenced in this study, rebuttal is difficult for students in that most of the students in the two groups were at levels 1 and 2 in the pre- and post-tests. Moreover, no more than half of the students achieved level 4 and no more than 20% of students reached level 5, regardless of claim, reason, evidence or rebuttal of the pre-test or the post-test. Obviously, to help more students achieve these two higher levels, a great and continuous effort is needed in the future.

RFS compresses the six argument components into four components which are differently arranged according to the specific assignment. It not only helped students understand the relationship between argument components but also helped them build meaningful connections. Since RFS is content relevant and designed with pedagogical considerations, such as the different characteristics of students reasoning, misunderstandings, and difficulties in constructing arguments, it allows students to concentrate on these issues when generating arguments for specific topics. Based on the results of this study, we would suggest that RFS can be adopted in chemistry classes to improve students' written arguments, especially students' ability in offering evidence and rebuttal. Besides, the SOLO taxonomy can be used as an analytical framework to evaluate the structure and quality of scientific arguments.

The present study was limited in several aspects. First, in our study, 10th grade high school students were involved, which did not allow us to generalize the results to other groups of students. Therefore, it is better to consider the findings of our research as exploratory and preliminary. Further research is needed to determine whether similar effects would be found in other student groups, such as university students or prospective teachers. Second, it is important to keep in mind that the results of this study might be different when the nature of the argumentation task changes. As shown in this study, RFS as a teaching strategy is domain specific, that is to say, it is varied dependent on the topic of scientific argumentation. Therefore, the results of the current study require further confirmation with different types of tasks to support a broader generalization. Finally, we used the SOLO taxonomy as a framework for assessing the quality of students' scientific arguments, but we did not compare our results with those of other studies. Future studies will explore whether there are differences in quality of arguments assessed by the SOLO taxonomy and other analytical frameworks.

## Conflicts of interest

There are no conflicts to declare.

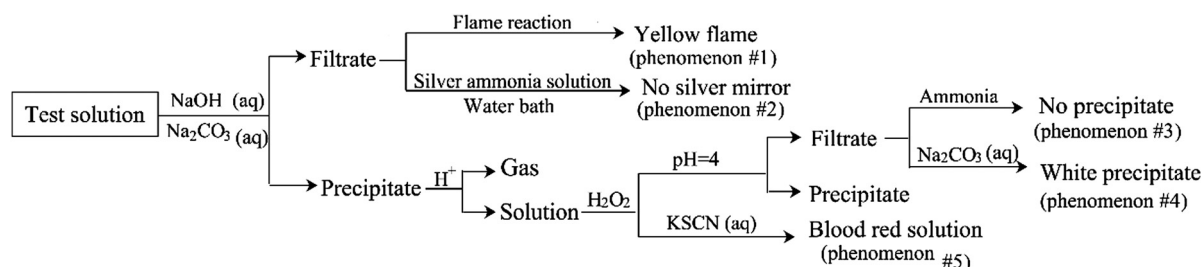
## Appendix 1. An example of written argument assignment

Part 1: Gluconate is a kind of food additive which is closely related to our lives and widely used in the food industry, such as nutritional supplements, curing agents, buffers, *etc.* In the laboratory, there is a bottle of gluconate compounds and its cations may be sodium, magnesium, iron and/or calcium. A flow chart for determining the cations in the gluconate test solution is shown as follows. Determine which cation(s) must be present in the test solution and explain the causes.

Note: (1) When pH of the solution is 4,  $\text{Fe}(\text{OH})_3$  is completely precipitated, but  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  do not precipitate.

(2) Ammonia is a weak alkaline and it can be partially ionized at room temperature. The reaction of ammonia with water is formulated as:  $\text{NH}_3 + \text{H}_2\text{O} \rightleftharpoons \text{NH}_3 \cdot \text{H}_2\text{O} \rightleftharpoons \text{NH}_4^+ + \text{OH}^-$ .

The flow chart for determining the cation(s) in the gluconate test solution is shown below:



Part 2 (for control group)

(1) Based on to the above information, please describe the experimental phenomena.

(2) Based on the above phenomena, what conclusions can you draw?

(3) Please explain how you came to the above conclusions.

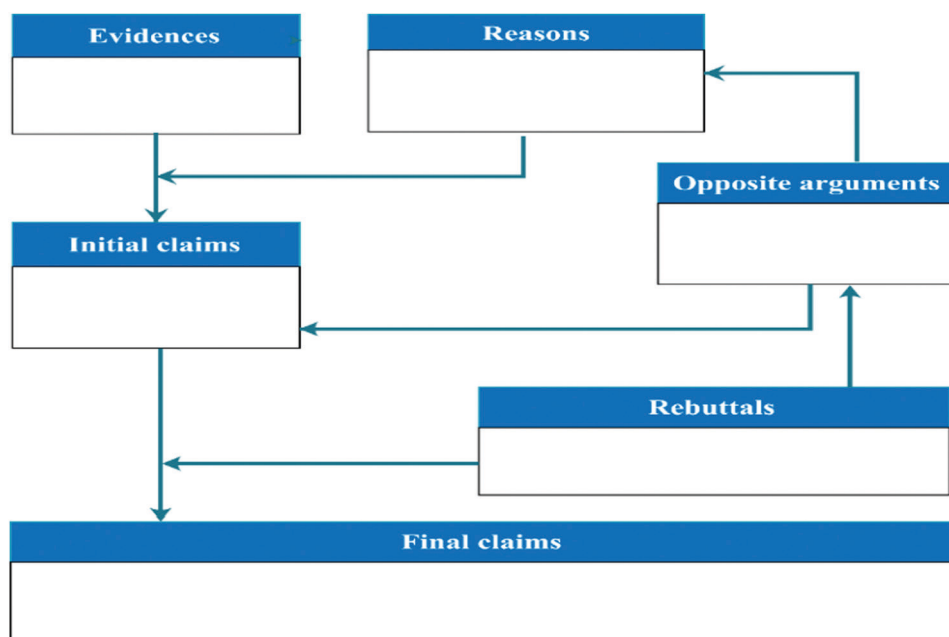
(4) If someone else provided an explanation that is inconsistent with yours and proposed the opposite argument based on this explanation, how would you raise rebuttal?

Part 2 (for experimental group)

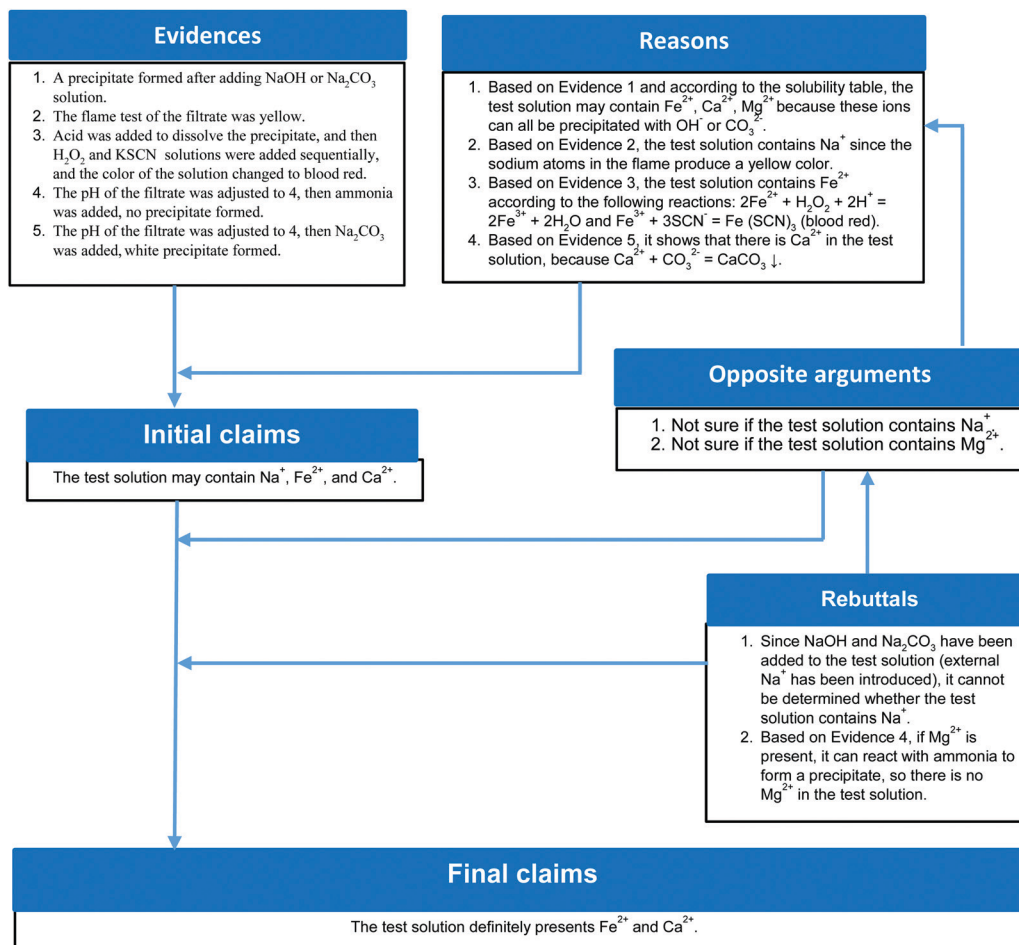
What is your claim? Use appropriate evidence and reason to support your claim.

What is the rebuttal that challenges the validity of the preliminary claim? Propose the rebuttal and final claim that you believe is the most valid or acceptable.

Write your argument in the corresponding blank of the RFS sheet.



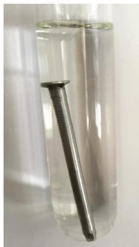





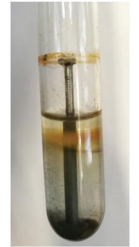
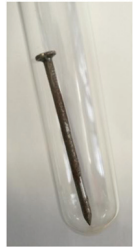


## Appendix 2. An example of written argument assignment by an experimental student



## Appendix 3. The measurement task of scientific argument

Class: _____	Name: _____
Age: _____	Gender: Boy ( ) Girl ( )

The nail is mainly composed of iron and contains a small amount of carbon. In humid air, it is affected by a variety of substances, and its surface is prone to rust. The color of rust is generally brownish red, but it would vary depending on the degree of corrosion. The appearance of rust is loose and porous, and the iron in the inner layer could be further corroded, resulting in it being gradually peeled off of its outer corrosion layer. In order to investigate 'the speed and influence factors of iron nail corrosion', five experiments were designed, of which cleaned iron nails were put in different external environments. Students conducted a 6 day experimental observation and recorded the phenomena, as shown in the table below.

Day	No.				
	1	2	3	4	5
Experimental conditions	Immersed in distilled water that cooled after boiling	Half of an iron nail immersed in NaCl solution	Naturally exposed to the air	Half of an iron nail immersed in distilled water	Immersed in distilled water cooled after boiling, and sealed with vegetable oil
The 1st day					
The 6th day					

- (1) According to the results in above table, please describe the experimental phenomena you observed.
- (2) According to the above phenomena, what conclusions can you draw?
- (3) Please explain how you came to the above conclusions?
- (4) If someone else provides other information that is inconsistent with yours and proposes the opposite argument based on these information, how would you raise rebuttal?

## Notes

This research was supported by the funding provided by the Guangdong Provincial Bureau of Education, and the Ministry of Education of China (DHA180440).

## Acknowledgements

Mr Sitong Chen (University of Macau) read the previous version of this article and gave invaluable comments and suggestion. We express our gratitude to him.

## References

- Aydeniz M., Pabuccu A., Cetin P. S. and Kaya E., (2012), Argumentation and students' conceptual understanding of properties and behaviors of gases, *Int. J. Sci. Math. Educ.*, **10**(6), 1303–1324.
- Bell, P. and Linn M. C., (2000), Scientific arguments as learning artifacts: Designing for learning from the web with KIE, *Int. J. Sci. Educ.*, **22**(8), 797–817.
- Belland B. R., Glazewski K. D. and Richardson J. C., (2008), A scaffolding framework to support the construction of evidence-based arguments among middle school students, *Educ. Tech. Res. Dev.*, **56**(4), 401–422.
- Berland L. K. and Reiser B. J., (2009), Making sense of argumentation and explanation. *Sci. Educ.*, **93**(1), 26–55.
- Bernholt S. and Parchmann I., (2011), Assessing the complexity of students' knowledge in chemistry, *Chem. Educ. Res. Pract.*, **12**(2), 167–173.
- Biggs J. B. and Collis K. F., (1982), *Evaluating the quality of learning: The SOLO taxonomy*, New York: Academic Press.
- Braaten M. and Windschitl M., (2011), Working toward a stronger conceptualization of scientific explanation for science education, *Sci. Educ.*, **95**(4), 639–669.
- Brown N. J. S., Furtak E.M., Timms M., Nagashima S. O. and Wilson M., (2010a), The evidence-based reasoning framework: Assessing scientific reasoning, *Educ. Assess.*, **15**(3), 123–141.
- Brown N. J. S., Nagashima S. O., Fu, A., Timms M. and Wilson M., (2010b), A framework for analyzing scientific reasoning in assessments, *Educ. Assess.*, **15**(3–4), 142–174.
- Caniglia J. C. and Meadows M., (2018), An application of the Solo taxonomy to classify strategies used by pre-service teachers to solve “one question problems”, *Aust. J. Teach. Educ.*, **43**(9), 75–89.
- Chang S. N. and Chiu M. H., (2008), Lakatos' scientific research programmes as a framework for analysing informal argumentation about socio-scientific issues, *Int. J. Sci. Educ.*, **30**(13), 1753–1773.
- Chin C. C., Yang W. C. and Tuan H. L., (2010), Exploring the impact of guided TAPPING scientific reading-writing activity on sixth graders. *Chin. J. Sci. Educ.*, **18**(5), 443–467 (in Chinese).

- Chin C. C., Yang W. C. and Tuan H. L., (2016), Argumentation in a socioscientific context and its influence on fundamental and derived science literacies. *Int. J. Sci. Math. Educ.*, **14**(4), 603–617.
- Claesgens J., Scalise K., Wilson M. and Stacy A., (2009), Mapping student understanding in chemistry: The perspectives of chemists, *Sci. Educ.*, **93**(1), 56–85.
- Clark D. B. and Sampson V., (2007), Personally-seeded discussions to scaffold online argumentation, *Int. J. Sci. Educ.*, **29**(3), 253–277.
- Cooper A. K. and Oliver-Hoyo M. T., (2016), Argument construction in understanding noncovalent interactions: a comparison of two argumentation frameworks, *Chem. Educ. Res. Pract.*, **17**(4), 1006–1018.
- Cornell R. M. and Schwertma U., (2003), *The Iron Oxides: Structure, properties, reactions, occurrence and uses*, 2nd edn, Weinheim: Wiley-VCH.
- Driver R., Newton P. and Osborne J. (2000), Establishing the norms of scientific argumentation in classrooms, *Sci. Educ.* **84**(3), 287–312.
- Erduran S., Simon S. and Osborne J. F., (2004), TAPPING into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse, *Sci. Educ.*, **88**(6), 915–933.
- European Union, (2006), Recommendation of the European Parliament on key competences for lifelong learning, Official Journal of the European Union, 30-12-2006, L 394/10–L 394/18.
- Han K. K., (2016), *Middle and high school students, scientific argumentation ability*, Doctoral thesis, Shanxi Normal University (in Chinese).
- Jordanou K. and Constantinou C. P., (2015), Supporting use of evidence in argumentation through practice in argumentation and reflection in the context of SOCRATES Learning environment, *Sci. Educ.*, **99**(2), 282–311.
- Jiménez-Aleixandre M. P. and Erduran S., (2007), Argumentation in Science Education: An overview, in Erduran S. and Jiménez-Aleixandre M. P. (ed.), *Argumentation in Science Education: Perspectives from classroom-based research*, Dordrecht: Springer, pp. 3–27.
- Katchevich D., Hofstein A. and Mamlok-Naaman R., (2013), Argumentation in the chemistry laboratory: Inquiry and confirmatory experiments, *Res. Sci. Educ.*, **43**(1), 317–345.
- McNeill K. L., (2011), Elementary students' views of explanation, argumentation and evidence and abilities to construct arguments over the school year, *J. Res. Sci. Teach.*, **48**(7), 793–823.
- McNeill K. L., Lizotte D. J., Krajcik J. and Marx R. W., (2006), Supporting students' construction of scientific explanations by fading scaffolds in instructional materials, *J. Learn. Sci.*, **15**(2), 153–191.
- Minogue, J., and Jones, G., (2009), Measuring the impact of haptic feedback using the SOLO taxonomy, *Int. J. Sci. Educ.*, **31**(10), 1359–1378.
- Moraes E. P., Confessor M. R. and Gasparotto L. H., (2015), Integrating mobile phones into science teaching to help students develop a procedure to evaluate the corrosion rate of iron in simulated seawater, *J. Chem. Educ.*, **92**(10), 1696–1699.
- National Research Council, (2012), *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*, Washington, DC: The National Academies Press.
- Novak A. M. and Treagust D. F., (2018), Adjusting claims as new evidence emerges: Do students incorporate new evidence into their scientific explanations? *J. Res. Sci. Teach.*, **55**(4), 526–549.
- Nussbaum E. M., (2002), Scaffolding argumentation in the social studies classroom, *Soc. Stud.*, **93**(2), 79–83.
- Oh S. and Jonassen D. H., (2007), Scaffolding online argumentation during problem solving, *J. Comput. Assisted Learn.*, **23**(2), 95–110.
- Olive J., (1991), Logo programming and geometric understanding: An in-depth study, *J. Res. Math. Educ.*, **22**(2), 90–111.
- Organisation for Economic Cooperation and Development (OECD), (2006), *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*, Paris: OECD Publishing.
- Organisation for Economic Cooperation and Development (OECD), (2013), *PISA 2015 Draft Science Framework*, Paris: OECD Publishing.
- Osborne J. F., Erduran S. and Simon S., (2004), Enhancing the quality of argument in school science, *J. Res. Sci. Teach.*, **41**(10), 994–1020.
- Potter M. K. and Kustra E., (2012), A primer on learning outcomes and the SOLO taxonomy, *Course Design for Constructive Alignment*, Winter, pp. 1–22.
- Ruiz-Primo M. A., Li M., Tsai S. P. and Schneider J., (2010), Testing one premise of scientific inquiry in science classrooms: Examining students' scientific explanations and student learning, *J. Res. Sci. Teach.*, **47**(5), 583–608.
- Ryu S. and Sandoval W. A., (2012), Improvements to elementary children's epistemic understanding from sustained argumentation, *Sci. Educ.*, **96**(3), 488–526.
- Sadler T. D. and Fowler S. R., (2006), A threshold model of content knowledge transfer for socioscientific argumentation. *Sci. Educ.*, **90**(6), 986–1004.
- Sampson V. and Clark D. B., (2008), Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions, *Sci. Educ.*, **92**(3), 447–472.
- Sampson V., Grooms J. and Walker J. P., (2011), Argument-driven inquiry as a way to help students learn how to participate in scientific argumentation and craft written arguments: An exploratory study, *Sci. Educ.*, **95**(2), 217–257.
- Sanders R. W., Crettol G. L., Brown J. D., Plummer P. T., Schendorf T. M., Oliphant A., Swithenbank S. B., Ferrante R. F. and Gray J. P., (2018), Teaching electrochemistry in the general chemistry laboratory through corrosion exercises, *J. Chem. Educ.*, **95**(5), 842–846.
- Sandoval W. A., (2003), Conceptual and epistemic aspects of students' scientific explanations, *J. Learn. Sci.*, **12**(1), 5–51.
- Saye J. W. and Brush T., (2002), Scaffolding critical reasoning about history and social issues in multimedia-supported learning environments, *Educ. Tech. Res. Dev.*, **50**(3), 77–96.
- Toulmin S. E., (1958), *The uses of argument*, Cambridge: Cambridge University Press.
- von Aufschnaiter C., Erduran S., Osborne J., and Simon S., (2008), Arguing to learn and learning to argue: Case studies

- of how students' argumentation relates to their scientific knowledge. *J. Res. Sci. Teach.*, **45**(1), 101–131.
- Wei B., (2019), Reconstructing school chemistry curriculum in the era of core competencies: A case from China. *J. Chem. Educ.*, **96**(7), 1359–1366.
- Wei B., (2020), The change in the intended Senior High School Chemistry Curriculum in China: focus on intellectual demands, *Chem. Educ. Res. Pract.*, **21**, 14–23.
- Wong V., Brophy J. and Dillon J., (2012), Combustion and redox reactions, in Taber K. S. (ed.), *Teaching secondary chemistry*, New 2nd, London: Association for Science Education, Hodder Education, pp. 199–252.
- Wood D., Bruner J. S. and Ross G., (1976), The role of tutoring in problem-solving, *J. Child Psychol. Psychiatry*, **17**(2), 89–100.
- Wright S. W., Stephen, W. W., (2005), Trusty or rusty? Oxidation rate of nails, *J. Chem. Educ.*, **82**(11), 1648A–1648B.
- Yeh K. H. and She H. C., (2010), On-line synchronous scientific argumentation learning: nurturing students' argumentation ability and conceptual change in science context, *Comput. Educ.*, **55**(2), 586–602.