This article was downloaded by: [Flinders University of South Australia]
On: 28 January 2015, At: 00:30
Publisher: Routledge
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

Click for updates

International Journal of Science Education
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tsed20

Scaffolding Middle School Students' Construction of Scientific Explanations: Comparing a cognitive versus a metacognitive evaluation approach

Chia-Yu Wang[a]
[a] Institute of Education, National Chiao Tung University, 1001 Ta-Hsueh Road, Hsinchu, Taiwan
Published online: 14 Nov 2014.

To cite this article: Chia-Yu Wang (2015) Scaffolding Middle School Students' Construction of Scientific Explanations: Comparing a cognitive versus a metacognitive evaluation approach, International Journal of Science Education, 37:2, 237-271, DOI: 10.1080/09500693.2014.979378

To link to this article: http://dx.doi.org/10.1080/09500693.2014.979378

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms &

Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

International Journal of Science Education, 2015
Vol. 37, No. 2, 237–271, http://dx.doi.org/10.1080/09500693.2014.979378

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Scaffolding Middle School Students' Construction of Scientific Explanations: Comparing a cognitive versus a metacognitive evaluation approach

Chia-Yu Wang*
Institute of Education, National Chiao Tung University, 1001 Ta-Hsueh Road, Hsinchu, Taiwan

This study investigated the effects of scaffolds as cognitive prompts and as metacognitive evaluation on seventh-grade students' growth of content knowledge and construction of scientific explanations in five inquiry-based biology activities. Students' scores on multiple-choice pretest and posttest and worksheets for five inquiry-based activities were analyzed. The results show that the students' content knowledge in all conditions significantly increased from the pretest to posttest. Incorporating cognitive prompts with the explanation scaffolds better facilitated knowledge integration and resulted in greater learning gains of content knowledge and better quality evidence and reasoning. The metacognitive evaluation instruction improved all explanation components, especially claims and reasoning. This metacognitive approach also significantly reduced students' over- or underestimation during peer-evaluation by refining their internal standards for the quality of scientific explanations. The ability to accurately evaluate the quality of explanations was strongly associated with better performance on explanation construction. The cognitive prompts and metacognitive evaluation instruction address different aspects of the challenges faced by the students, and show different effects on the enhancement of content knowledge and the quality of scientific explanations. Future directions and suggestions are provided for improving the design of the scaffolds to facilitate the construction of scientific explanations.

Keywords: *Biology; Cognitive prompts; Metacognitive evaluation; Scaffolding; Scientific explanation*

*Email: cwg25@mail.nctu.edu.tw

© 2014 Taylor & Francis

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Introduction

The standard documents of the USA for K-12 and post-secondary education (College Board, 2009; National Research Council [NRC], 2011) and the Programme for International Student Assessment (PISA) 2015 Draft Science Framework (OECD, 2013) stress the importance of engaging students in scientific practice. One of the central goals of science education is to prepare students to synthesize and evaluate scientific explanations. These standard documents and the PISA framework explicitly state the learning goals of developing students' abilities to construct and interpret evidence-based explanations and models of the natural world, and to evaluate their own or others' explanations by judging logical or flawed connections between evidence and conclusions. Immersing students in scientific inquiry and providing them with experience in the construction of sound scientific explanations may change their epistemic views of science (McNeill, Lizotte, Krajcik, & Marx, 2006). The ability to construct scientific explanations is also a springboard for developing other abilities, such as identifying and arguing against weaknesses in scientific arguments or evaluating models based on evidence and scientific knowledge.

Developing proficiency in constructing scientific explanations is a complex and cognitively demanding process. Students must possess content knowledge, know the key features that constitute appropriate explanations, and understand how the constructed explanation relates to knowledge-building practices. Previous research has shown that students encounter various obstacles, and attempts have been made to help them overcome some of these challenges (e.g. Chin & Osborne, 2010; McNeill et al., 2006; Sampson, Grooms, & Walker, 2011). Based on analyses of the nature of explanation and scientific practices, some instructional frameworks have been proposed which holistically incorporate scaffolds to support epistemic, structural, conceptual, and/or social aspects of scientific explanation. For example, the studies of McNeill and colleagues (McNeill & Krajcik, 2008; McNeill et al., 2006) have shown that, with the mediation of a holistic instructional model, middle school students have made substantial gains in their competency of synthesizing explanations; however, the students' average scores for evidence and reasoning remained below half of the maximum scores after interventions. These findings indicate that many students were still not proficient and required further supports to reach a competent level. The limited effects of these studies suggest the existence of other cognitive or metacognitive obstacles that must be diagnosed and addressed before students' competency can be further improved.

Previous studies have indicated a close relationship between the quality of students' conceptual understanding and their ability to synthesize quality explanations (e.g. Chin & Osborne, 2010; McNeill et al., 2006). Cognitive scaffolds such as using prompts to highlight key principles (Sandoval, 2003) or prompting students to post questions for deep-thinking processes (Chin & Osborne, 2010) have been utilized to enhance the quality of explanations. However, these studies examined the impacts of an instructional framework as a whole; the effects of the scaffolds that address different learning obstacles (such as supplying conceptual resources or using external criteria for evaluation) were not differentiated. Whether incorporating cognitive or metacognitive scaffolds affects the quality of the explanations and conceptual understanding, as well as which aspect of the explanations benefits remain unclear. Although Chin and Osborne (2010) and Sampson et al. (2011) have emphasized the importance of introducing criteria of quality explanation and encouraging students to use the criteria for evaluation, what obstacles students encounter during evaluation as well as whether and how this type of metacognitive scaffolding influences the learning of explanation are overlooked.

While several instructional models have been developed to cultivate competency of scientific explanation in the classroom, adding specific cognitive or metacognitive scaffolds may further benefit learning outcomes. Thus, the present study aims at designing two types of scaffolds, a cognitive and a metacognitive approach, and to examine their effects on improving middle school students' scientific explanations. To address this aim, this study adopted a finer-grained lens to analyze the cognitive and metacognitive obstacles which potentially occur during the processes of explanation construction and evaluation, in addition to those difficulties reported in the current scientific explanation literature. The literature regarding effective designs for fostering cognitive reasoning and adequate metacognitive evaluation was also reviewed to synthesize guidelines in order to design corresponding scaffolds in the present study. To understand whether and how the cognitive and metacognitive scaffolds attribute to the learning of explanation in addition to the original instructional framework, a quasi-experimental design was used for examining the effects of the two scaffolding designs in a separate manner. In addition, trends for change were reported to provide information regarding whether the cognitive and metacognitive scaffolds affect the quality of claims, evidence, and reasoning differently, and if the changing process supported by either form of scaffolding is evolutionary or rather revolutionary. The information is valuable for researchers to determine the duration of the scaffolds and the most appropriate timing for fading.

One point worth noting is that the aim of the present study was not to propose a new instructional framework for fostering scientific explanation. Rather, the author seeks to adopt a widely implemented, existing model as a base on which to build additional scaffolds. McNeill et al.'s framework (McNeill & Krajcik, 2008; McNeill et al., 2006) was selected for the following reasons: (1) it has embedded scaffolds specifically developed to address students' difficulties regarding explanation construction; (2) the designs of their scaffolds and the rationales behind them, criteria for evaluating the quality of students' explanations, as well as the context and sequence of implementations are explicitly delineated; (3) in addition, their framework and curriculum have been implemented with a large group of middle school students, and both the effective and not-efficient aspects of instructional implementation have been reported (see McNeill & Krajcik, 2008), which allows other researchers to duplicate their designs successfully.

Definition of Scientific Explanation

Scientific explanations frame the goal of inquiry as making sense of a phenomenon based on current scientific knowledge, and articulating and convincing others of

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

that understanding (Sandoval & Reiser, 2004). Science educators generally accept that explanations are attempts to provide an explanatory account of natural phenomena that specifies what happened and/or why it occurred (e.g. Berland & Reiser, 2009; McNeill et al., 2006; Osborne & Patterson, 2011). However, viewpoints concerning what counts as explanations and what is involved in the construction of scientific explanations vary. McNeill et al. (2006), for instance, did not differentiate explanations from arguments, but considered that explanations can be constructed for two goals: to explain a natural phenomenon or to support an individual's opinion or belief. Thus, their descriptions of claim, evidence, and reasoning in many ways are like the elements of arguments found in instructional models of argumentation (e.g. Sampson et al., 2011). Osborne and Patterson (2011), on the other hand, examined the nature and purpose of the practices, thinking that arguments should be distinguished from explanations because they are driven by different purposes. The purpose for forming explanations is to make sense of a phenomenon based on known knowledge; thus, explanations are neither constructed out of data nor need to be justified (Osborne & Patterson, 2011). The purpose of constructing arguments is to persuade; thus, arguments seek to determine whether an explanation is better than competing accounts by evaluating the extent to which the claim, data, and justification of an explanation are sufficient and coherent (Osborne & Patterson, 2011).

Agreeing with Osborne and Patterson (2011), Berland and Reiser (2009) took a pedagogical stance to examine the goals of instructional models that emphasize scientific explanations (e.g. McNeill et al., 2006; Sandoval & Reiser, 2004) or argumentation (e.g. Osborne, Erduran, & Simon, 2004; Sampson et al., 2011). Constructing and defining scientific explanations were then combined into a single practice because these instructional models often share similar instructional goals. These goals include helping students: (1) to use evidence to make sense of a phenomenon, (2) to articulate explanations by explicitly connecting the evidence to their claim using existing knowledge, and/or (3) to persuade others through reconciling competing ideas in social discourses to derive the most robust explanation. Moreover, Braaten and Windschitl (2011) took a typological stance to analyze different types of explanations. Their definitions of explanations were extended to include (1) statements for explication (e.g. stating definition of terminology), (2) a cause–effect relation account proposed to explain a phenomenon, or (3) an argument justified with evidence and reasons.

This study adopted McNeill et al.'s (2006) descriptions, depicting scientific explanation as constituting three major components, specifically claims, evidence, and reasoning (warrant and backing). Constructing scientific explanation involves generating claims that account for a phenomenon by connecting scientific principles with the evidence at hand and justifying claims using appropriate evidence and scientific ideas. The instructional goals emphasized in McNeill et al.'s framework regarding making sense of natural phenomena and gaining competency of constructing adequate explanations align with the former two goals of Berland and Reiser (2009), whereas their definitions of explanations fall in the latter two types of Braaten and Windschitl's (2011) categorization. Sampson and Clark (2008) have suggested

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

three perspectives, including the structure of an explanation, conceptual quality, and epistemic quality to characterize criteria for quality explanation. McNeill et al. (2006) depict a quality scientific explanation which uses both evidence and reasoning appropriately and sufficiently to support the claim. McNeill et al.'s criteria were then analyzed according to Sampson and Clark's characterization. In terms of its structure, an explanation is depicted as a claim that consists of supporting evidence and reasoning for justification. Their criteria assess the conceptual quality of an explanation by measuring the extent to which students can articulate a causal claim in a specific science topic and support that claim with appropriate data and scientific principles. The epistemic quality of an explanation was evaluated by assessing the extent to which the data and scientific principles used by the students to support the claim are coherent and sufficient.

Challenges in Constructing Scientific Explanations and Possible Causes

Generating and crafting written explanations are challenging for many students. The quality of scientific explanation may be affected by students' limited understanding of scientific explanations or of how to write such explanations. Previous studies have revealed that middle school students' explanations are dominated by claims (McNeill et al., 2006). Students may draw on personal views or other knowledge to explain a phenomenon rather than using the evidence at hand (McNeill et al., 2006; Sandoval & Reiser, 2004). Some students tended to merely describe what they observed rather than providing a causal account of why and how a phenomenon happened (Sampson et al., 2011). Some struggle to understand what qualifies as evidence, while others may draw data from an investigation but have difficulties in determining whether the evidence is appropriate or sufficient to support their claim (McNeill et al., 2006; Sandoval & Millwood, 2005). Many students offer ambiguous statements in which evidence is not clearly distinguished from inferences (Berland & Reiser, 2009). Some of these obstacles may be attributed to students either not understanding the relationships among forming a hypothesis, collecting data, and drawing conclusions during scientific inquiry, or not knowing the essential components that constitute a scientific explanation (McNeill et al., 2006; Sampson et al., 2011).

Even when they possess content knowledge, students may have a difficult time using appropriate scientific principles to justify their reasons for selecting evidence to support their claim (Sandoval & Millwood, 2005). Some of these difficulties may be attributed to students' flawed cognitive reasoning. Deficient strategic knowledge or procedural difficulties that cause flawed cognitive reasoning have been described as functional fixedness and functional reduction (Furió, Calatayud, Bárcenas, & Padilla, 2000). Functional fixedness describes the situation in which students infer or support an idea based on common sense evidence without considering scientific knowledge. Predictions or explanations synthesized from common sense reasoning often synthesize answers differing from those held by scientists. Functional reduction indicates students' tendency to reduce the intrinsic complexity of a problem during their reasoning process (Furió et al., 2000) so as to reduce the information-processing load. In cases of functional reduction, students may focus on only one or a few factors or variables and ignore others when constructing an explanation that, in fact, requires the consideration of more factors. For example, students intuitively inferred that the more electrons in an atom, the larger it is, without considering the existence of electrostatic force between the positive nucleus and the negative electrons (Talanquer, 2006). Furió et al.'s study showed that high school and undergraduate students tend to be hindered by functional fixedness and functional reduction when solving complex chemistry problems which involve simultaneously considering several conditions or which require multiple reasoning steps (e.g. determining molecular polarity by considering not only the molecular shape but also the electronegativity of bonded atoms). Talanquer (2006) also identified several alternative conceptions in chemistry resulting from the influences of the flawed reasoning. It is possible that the learning of some middle school students is impaired by functional fixedness and functional reduction when forming, justifying, and evaluating explanations.

In some instances, the effectiveness of instruction that addresses these obstacles may be constrained because students do not always spontaneously monitor their own learning or may misjudge, mostly overestimate, their performance or ability during self-evaluation. Previous studies have shown that students developed a false sense of competency, rating themselves as more knowledgeable of what they had understood, if no criteria were provided, when they were asked to explain what they had learned from a set of definitions in biology (Lipko et al., 2009) or from watching visualizations in chemistry (Chiu & Linn, 2012). Studies such as that of Davis (2000) have shown that students who received reflective prompts and external criteria for evaluating the quality of their own project demonstrated greater understanding of scientific practices. However, some students may have limited ability to evaluate the quality of their performance, even if external criteria are available for comparison. For instance, in one study, when students were provided with feedback on their performance over multiple tests, only high-achieving students benefited from the external feedback and increased the accuracy of their self-evaluations, while the low-achieving students had difficulties recognizing inconsistencies between their performance and the external criteria (Rawson & Dunlosky, 2007). Adequate self-evaluation would help learners to identify components that need to be improved and generate a plan for improvement in the next learning cycle; inadequately evaluating the quality of one's own performance may hinder one's learning (Winne & Hadwin, 1998).

Potential Scaffolding Designs to Address the Challenges of Explanation Construction

Ample studies have developed instructional models (e.g. Chin & Osborne, 2010; McNeill & Krajcik, 2008; McNeill et al., 2006; Sampson et al., 2011) to promote scientific explanation and argumentation in science classrooms. Several instructional supports (scaffolds) are proposed to enhance students' epistemic and structural aspects of explanations in these models. Scaffolds that aim at enhancing the epistemic quality include specifying the goal for explanations as making sense of phenomena (e.g. McNeill et al., 2006) or persuasion (e.g. Sampson et al., 2011). Specifying the

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

goals of these practices will help learners see the need to support or validate a claim using appropriate evidence and reasoning. This scaffold is often accompanied with explicitly introducing attributes that characterize what counts as good explanations (Braaten & Windschitl, 2011; Sampson et al., 2011) by introducing specific criteria (Chin & Osborne, 2010; Sampson et al., 2011) or explaining with both strong and weak examples (Sampson et al., 2011). Once the students have gained an initial understanding of the goals and the nature of explanations, other scaffolds then come into play to help students articulate their explanations. These scaffolds include helping students differentiate claim, evidence, and supports for inferences (reasoning) into three components (e.g. McNeill et al., 2006; Sampson et al., 2011) and modeling the construction of explanations (McNeill et al., 2006; Sandoval & Reiser, 2004). Supplementing visual supports (e.g. an argument diagram (Chin & Osborne, 2010) or a written scaffold (McNeill et al., 2006)) that illustrate the relations between components of explanations will help the learners focus and organize their explanations. Students also require multiple practices to articulate their explanations both orally and visually (Berland & Reiser, 2009; Chin & Osborne, 2010; McNeill et al., 2006; Sampson et al., 2011). In the meantime, encouraging students to use appropriate standards for what counts as quality explanations to evaluate some given examples or to determine the quality of their own or peers' artifacts will help students to be more metacognitive as they work on instructional activities (Sampson et al., 2011). There are also scaffolds aimed at supporting social and collaborative discourses by having students share their arguments and critique the work of others (Sampson et al., 2011) or having them pose questions and discuss opposite viewpoints in small groups (Chin & Osborne, 2010). These social scaffolds are to support establishing the structure and social norm of peer interactions and to create a community of learners that value evidence and critical thinking (Sampson et al., 2011).

Relatively few instructional models incorporate scaffolds that facilitate cognitive reasoning, and very few specifically take metacognition into consideration. Among the scaffolds designed to support cognitive reasoning, prompts are one of the most common strategies. Prompts are questions or elicitations that aim to stimulate active thinking or elicit learning strategies such as self-explanation (Schworm & Renkl, 2007). For example, Chin and Osborne (2010) supported students posting questions in small groups to foster active thinking and engage students in scientific practices. To resolve the obstacle of flawed cognitive reasoning, prompts can be used to support cognitive structuring by asking questions that require an active cognitive answer or giving hints without supplying the entire solution (Schworm & Renkl, 2007). Prompts which provide conceptual resources (e.g. evidence statements or relevant principles) that students can draw on can also facilitate inference generation (Chin & Osborne, 2010). These cognitive prompts, which are grounded in domain-specific knowledge, may help students activate existing knowledge and identify salient features in the content (Bulu & Pedersen, 2010). In addition, prompts can be used to guide students to determine scientific principles that are appropriate for developing justifications for their reasoning (Bulu & Pedersen, 2010; Sandoval,

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

2003). These justifications may otherwise be overlooked because of functional fixedness or functional reduction. Previous studies have shown that providing students with prompts that highlight the causal components of important domain theories helps students produce explanations for inquiry (Sandoval, 2003). In addition, scaffolds that provide content-specific guidance facilitate the learning of science content and problem representation better than general problem-solving scaffolds (Bulu & Pedersen, 2010). Conceptual knowledge was particularly fostered when students were prompted to self-explain their rationale of applying particular principles (Berthold, Eysink, & Renkl, 2009). More elaborate, higher quality arguments could be derived as well when students are prompted to ask questions that address substantive scientific concepts (Chin & Osborne, 2010).

To counteract inadequate self-evaluation, specifically overconfidence, a common scaffolding strategy is to provide students with feedback regarding their own learning outcomes or offer opportunities to self- or peer-evaluate using external criteria. Outcome feedback leads to conservative judgments (Rawson & Dunlosky, 2007), whereas peer-evaluation may enhance student assessors' understanding of the criteria for the quality of artifacts which they may later apply to their own tasks; it may also motivate self-monitoring, and influence the planning of learning strategies for the subsequent learning cycle (Chang, Quintana, & Krajcik, 2010). Specific guidance, such as creating a context in which using the criteria for evaluation is valuable and makes sense (Sampson et al., 2011), is needed as well to help students adapt to the external criteria for evaluation. Previous studies have shown that students who receive feedback or experience peer-evaluation are more accurate in their self-judgments of model evaluation and outperform their counterparts (Chang et al., 2010). However, even with the presence of external criteria, some students may have difficulty in recognizing inconsistencies between their performance and the criteria. The amount of information conveyed in different designs of external criteria affects the adequacy of the student evaluation. For instance, Lipko et al. (2009) investigated middle school students' skills of judging their learning of key term definitions in biology and whether their overconfidence regarding their ability to self-evaluate could be reduced by providing different standards. When the complete, correct definitions were provided as standards (full-definition standards), students' overconfidence was reduced; however, the influence was selective: the standards helped the students identify incorrect responses but had a minimal influence on their ability to distinguish partially correct from completely correct responses. This may be attributed to the full-definition standards not providing sufficient cues to help the students recognize the specific information required for a response to be considered correct and the important components missing in their responses. Overconfidence was further reduced when key components of each definition were highlighted (idea-unit standards), and the students had to indicate whether each key component was included in their response when using the idea-unit standards for self-evaluation. Discussing strong and weak examples with corresponding criteria may also help students better interpret external criteria (Osborne et al., 2004).

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

In the present study, McNeill et al.'s (2006) instructional framework was adopted to address the first obstacle of students' limited understanding of scientific explanation because their framework has embedded scaffolds for enhancing the epistemic and structural quality of explanations. Other scaffolds such as evaluating some given examples using external criteria were added for enrichment. In this study, scientific explanation was introduced in the context of a scientific investigation during a focal lesson to help students obtain an initial understanding regarding the purpose and practice of constructing explanations. The goal of the investigations was framed so as to make sense of natural phenomena. The teachers then introduced the components and structure of explanations and depicted what counts as good explanations by specifying the relations among claims, evidence, and reasoning. They also modeled how to construct an explanation, and supplemented the instruction with McNeill et al.'s written scaffold. The written explanation scaffold was designed to help students understand the general structure and components of a scientific explanation and can later be used during explanation construction to remind them of the key features of an explanation. It consists of reminders that a claim is a sentence that states one's conclusion about the question to be answered. Students are also prompted to write evidence for the question to be answered, such as 'provide two pieces of data that support your claim', and reasoning, such as 'write a statement that connects your evidence to your claim'. The teacher also critiqued multiple examples of explanations to help students learn how to adapt the general explanation framework for a specific context. After the focal lesson, the students had opportunities to discuss and articulate, both orally and verbally, their explanation in each unit.

McNeill et al.'s (2006) framework and scaffolding designs were then used as a base on which to build additional cognitive or metacognitive scaffolds. To design cognitive scaffolds, a panel consisting of a science educator, a science education graduate student, and two participating teachers, each with at least five years' experience of teaching biology, identified important scientific principles or theories for explaining each phenomenon under study. Cognitive prompts were then developed in a question format and tailored for each inquiry activity to direct students' attention specifically to consider the salient knowledge of that particular phenomenon. In contrast to Chin and Osborne's (2010) design, in which the students were prompted to pose various questions, the prompts used in this study provide conceptual resources through directing students to discuss and respond to the given questions using related scientific knowledge learned in the classes. A hint was also placed in the reasoning section of the worksheet to remind the students to draw on their responses to those cognitive questions as resources to connect evidence with their claim.

The design of the metacognitive scaffolds comprises several features that aim at enhancing adequate evaluation, specifically counteracting overestimation. The criteria for quality explanation were revised from the full-definition format to the idea-unit format to highlight specific information required for a response to be considered correct or partially correct. The teachers then guided the students to discuss and indicate what components were missing when using the idea-unit standards to evaluate weak examples as well as to revise them into correct ones. At the end of the

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

investigation, students were also asked to use the standards to evaluate the quality of others' explanations. The opportunity for peer-evaluation creates a need to understand the content of the standards and to be better at explanation evaluation. The aforementioned metacognitive scaffolds were rehearsed in each unit. Ford and Yore (2012) proposed that metacognition involves thinking about your learning as you are learning to improve your learning. Through immersing students in the cyclic processes of metacognitive evaluation, the students are supported in examining their own knowledge and the quality of their work of scientific explanation, overseeing their process of explanation construction, as well as in thinking about how to improve their own learning in the subsequent units.

The present study builds on McNeill et al.'s (2006) efforts regarding their explanation framework with the scaffolding designs (the E-only group) and to either incorporate cognitive prompts to address flawed cognitive reasoning (the C + E group) or combine metacognitive evaluation instruction to resolve inadequate self-evaluation using researcher-developed idea-unit standards during peer-evaluation (the M + E group). The aims and designs of the two scaffolds differ, and therefore may result in different learning outcomes in terms of gains in conceptual knowledge and/or quality of explanations. Thus, although some researchers may argue for combining cognitive and metacognitive approaches to improve learning outcomes, the present study seeks to understand the different effects of scaffolding designs using cognitive and metacognitive evaluation approaches. Moreover, previous studies investigating the effects of cognitive or self-evaluation approaches tested them on short-term interventions (e.g. Lipko et al., 2009; Sandoval, 2003) or did not report changes of learning over time (e.g. Chang et al., 2010; Sampson et al., 2011). The trajectory of middle school students' progress toward proficiency in explanation construction requires further investigation. Specifically, this study investigated the following questions:

(1) How do different scaffolding designs (C + E, M + E, and E-only) influence students' learning outcomes of biology concepts?
(2) How do different scaffolding designs (C + E, M + E, and E-only) influence students' quality of scientific explanation (claims, evidence, and reasoning) over units?
(3) For students who receive metacognitive scaffolds (M + E), how is their quality of scientific explanation (claims, evidence, and reasoning) affected by their level of evaluation adequacy?

Methods

Subjects

This study utilized a quasi-experimental design. A total of 173 seventh-grade students from 6 average classes at 2 middle schools in northern Taiwan participated. Two biology teachers taught three classes each, which were randomly assigned into one of the three conditions: C + E, M + E, and E-only. All students and the participating teachers were native Mandarin speakers. Thus, the written materials, scaffolds, and two instruments (described later) were implemented in Chinese. Students' written responses were also collected and analyzed in Chinese.

Course Context

At the beginning of the experimental instruction period, students in all instructional modes received a focal lesson that helped them understand the logic behind scientific inquiry practice and the rationale behind scientific explanation. Definitions of the components of scientific explanation, adapted from McNeill et al. (2006), were introduced and discussed with all students to explain what qualifies as high-quality claims, evidence, and reasoning using clear and complete examples. Five inquiry-based biology units on the topics of cell osmosis, enzymes, transpiration, response time, and plant reproduction were then implemented in sequence. The focus of the inquiry activities for all three groups was to engage the students in synthesizing a hypothesis, collecting and analyzing data or interpreting second-hand data, and constructing a scientific explanation while exploring a natural phenomenon. The students worked in small groups to collaboratively complete the investigation, analyze the data, and discuss the phenomena; however, they wrote their own explanations on the investigation sheet. McNeill et al.'s written explanation scaffold was used as the generic support to assist students in all groups with writing explanations.

The class activities and instructional sequence for each unit for the three groups are presented in Table 1. The focal lesson took one class period, and each of the five units (including stages 1–3) was completed within three class periods in a week, with one week between units. The total duration of the instruction for all three groups was 10 weeks. The design and implementation of the focal lesson and the five inquiry activities (stage 2) were the same for all three groups. The only differences among the three instructional modes lay in the characteristics of the scaffolding design concerning facilitating scientific explanation in stages 1 and 3, as illustrated as follows:

In stage 1 of each unit, students of the E-only group were introduced to the criteria of a high-quality scientific explanation (full-definition standards), adapted from McNeill et al.'s (2006) definitions of claims, evidence, and reasoning. Examples of complete statements for claims, evidence, and reasoning were provided, and the students were guided to discuss whether and why they considered the examples as good explanations. After each inquiry activity, the E-only group formed their scientific explanations using the written explanation scaffold (stage 3).

Students of the C + E group received the same scaffold design in stage 1. The difference between this group and the E-only group was that cognitive prompts were also used to support the students' cognitive thinking in a specific content area when they formed explanations in stage 3. Prior to forming explanations, prompts were provided as written questions to highlight the salient content knowledge related to the topic. In addition, hints were provided to help the students develop sound reasoning associated with the related knowledge activated by the prompting

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Table 1. Sequence of class activities and major differences among the three instructional modes

	Major differences		
Class activity	E-only ($n = 55$)	C + E ($n = 59$)	M + E ($n = 59$)
Focal lesson	No scaffolds	[a]	[a]
For units 1–5			
Stage 1: Reiterate the rationale and discuss criteria for a high-quality scientific explanation	Discuss using full-definition standards as the criteria Discuss using only the example of complete statements	[a]	Discuss using idea-unit standards as the criteria Discuss with examples of complete, partial, and incomplete statements
Stage 2: Experience inquiry in small groups	Observe a natural phenomenon Discuss and form a hypothesis Conduct an investigation to collect data or interpret second-hand data	[a]	[a]
Stage 3: Form a scientific explanation		Respond to cognitive prompt questions	
	Form an explanation using the written explanation scaffold adapted from McNeil et al. (2006)	[a]	[a]
		Students were reminded to associate their responses to the cognitive prompts with their reasoning statements	Evaluate peer's quality of explanation using idea-unit standards The rated explanation was returned to its owner who decided whether he/she agreed with the evaluation

[a]Same as the E-only group.

questions. Appendix 1 provides an example of explanation scaffolding for students' investigation sheets of the three instructional modes. The topics investigated, the cognitive question prompts, and the key conceptions identified in the five units are illustrated in Table 2. For an illustrative purpose, the examples presented here and in the following sections were first translated by a science educator into English. Both the original and the translated versions were then reviewed by a biologist who is fluent in both languages. Discrepancies between the two versions were discussed and resolved by the two experts through rounds of discussion to ensure the Chinese and English versions of the examples were conceptually equivalent.

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Table 2. The topic investigated, the cognitive question prompts, and key conceptions

The topic investigated	Cognitive question prompts used in the C + E group	Key conceptions
(1) *Cell osmosis*: Will watering a plant using seawater kill the plant or help it grow?	Compare the concentration of total dissolved solutes in the seawater and in the cytosol of the plant's roots. Which one is higher? How does the gradient of concentrations influence the movement of water inside or outside the plant cells?	• Cell membrane is selectively permeable. Water and oxygen move freely across the cell membrane by diffusion, but solutes (e.g. salt) cannot • Osmosis describes the diffusion of water through the cell membrane from a more dilute solution to a more concentrated solution • When cells are placed in a less concentrated solution (e.g. distilled water), there is less solute outside the cells. Thus, water will move into the cells. Cells will swell and may burst. When cells are placed in a more concentrated solution (e.g. seawater), there is more solute outside the cells. Water will move outward from the cells, and the cells will shrink and shrivel. Water will move through the membrane from both sides at equal rates when the concentration is the same on both sides of the membrane

(*Continued*)

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Table 2. Continued

The topic investigated	Cognitive question prompts used in the C + E group	Key conceptions
(2) *Enzymes*: How does temperature influence level of activity for amylase (the length of time needed for amylolysis)?	Under what temperature are amylases most activated? How do you know?	• Most enzymes are very specific for a certain substrate. For instance, amylase is the enzyme that breaks down starch into sugar maltose. Amylase can be found in human saliva and the pancreas • Enzymes, including amylase, are proteins. Exposure to high heat will denature proteins. If denatured, amylase can no longer act as a catalyst for amylolysis • The temperature at which an enzyme works best (the optimum temperature) varies for different types of enzymes. Increasing the temperature will speed up the reaction until the optimum temperature is reached, but over that temperature the amylases will start to denature and the reaction rate of amylolysis will decrease • Benedict's solution reacts with reducing sugars like maltose but not with starch. Obtaining a green resulting solution indicates the presence of only a little reducing sugar. Getting a color from yellow, orange, to red indicates the presence of reducing sugars from a little bit more to a lot

(3) *Transpiration*: How does the number of leaves affect the speed of transpiration?	Where does water leave a plant during transpiration? (What is the evidence?)	• Transpiration is the process by which moisture is carried through plants from roots to small pores (so called stomata) on the underside of leaves
	How does the water level of the experimental group (a plant with more leaves) differ from that of the control group (a plant with fewer leaves)? What drives plant transpiration?	• Transpiration is driven by the evaporation of water from the stomata into the atmosphere. You can tape strips of cobalt chloride test paper beneath the plant leaves and observe if moisture escaping from the stomata turns the blue cobalt chloride test papers pink • In the same condition, the transpiration rate increases when the number of leaves and of stomata increases. Other factors may affect the transpiration rate, including temperature, humidity, and sunlight intensity
(4) *Response time*: Perform a ruler drop test. Catch the ruler as soon as it falls and record the level at which you catch it as an indicator of your response time. Is it possible to reduce your reaction time to zero seconds by performing this experiment repeatedly?	How is the length of the neuropath (from your eyes to your hand) for passing the electrical signal related to the length of the response time?	• The nervous system mediates communication between different parts of the body. The nervous system consists of the brain, the spinal cord, and all the nerves that deliver signals to the spinal cord
	Can the length of the neuropath be shortened through performing the reaction time experiment repeatedly?	• When you see your partner drop the ruler (a stimulus), sensory neurons in your eyes are activated and relay a message to the brain. The brain then processes the message, makes a decision, and sends out a signal that travels to the spinal cord and is finally delivered to the motor neurons, telling the muscles in your fingers to catch the ruler (the body's response) • Reaction time is the time taken between the application of a stimulus and the body's response to the stimulus. Reaction time can be improved with practice, up to a point, but your body does not react instantly

(*Continued*)

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Table 2. Continued

The topic investigated	Cognitive question prompts used in the C + E group	Key conceptions
(5) *Plant reproduction:* Mimi has two strawberry plants. Which approach (reproduce by runner propagation or pollinate a mother plant and then reproduce with seeds) should she choose to get sweeter strawberries?	Sexual or asexual reproduction, which one involves haploid gametes resulting from meiosis? Why does fertilization produce filial generation with more diverse characteristics?	• Plant reproduction is the production of offspring, which can be accomplished by sexual or asexual reproduction • Sexual reproduction produces offspring by the fusion of gametes. Sexual reproduction results in offspring genetically different from the parent(s) and increases genetic diversity • Sexual reproduction involves processes of meiosis and fertilization. Meiosis refers to reduction division because the genetic complement of each daughter cell is reduced by half. During fertilization, a male gamete fuses with a female gamete, and the chromosome is restored to a complete diploid number • Asexual reproduction involves mitosis and produces offspring without the fusion of gametes. Mitosis refers to cell division that produces daughter cells with the genetic complement identical to the mother cell. Thus, the offspring will be exact genetic copies of the parent individual except for mutation

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

The goal of the scaffolding design for the M + E group was to address students' inadequate self-evaluation by improving their understanding of quality scientific explanation using a set of refined criteria (idea-unit standards), and gradually internalizing the criteria through cyclical peer-evaluations. Thus, the definition of claims in the full-definition standards, for example, was refined into idea-unit standards by additionally specifying two key components: (1) write a complete sentence (rather than merely giving a simple phrase or a few terms), and (2) provide a causal account of the investigated phenomenon (e.g. how the independent variable affects the dependent variable) to help the students recognize what specific information is required for a high-quality claim. In stage 1 of each unit, idea-unit standards were reintroduced. Examples of incomplete, partial, and complete statements for claim, evidence, and reasoning were also provided for the students to discuss in pairs. The students applied the idea-unit standards to evaluate and revise partial and incomplete examples. After their explanations were written in stage 3, the idea-unit standards were reiterated and supplemented with domain-specific guidance to depict how the general idea-unit standards were applied to a domain-specific evaluation. For instance, when evaluating the claim for the unit on transpiration, the claim definitions of the idea-unit standards were illustrated with guidance such as 'describe the relationship between the number of leaves and the speed of transpiration'. The students then worked in pairs to rate their partners' explanations as incomplete, partial, or complete based on the criteria. The rated explanations were then returned to their owners who decided whether they agreed with them. The differences between the design of the full-definition and the idea-unit standards are illustrated in Table 3. Notably, students of all instructional modes were supported in writing their explanations by written scaffolds on their investigation sheets and by other social supports, such as their peers and teacher.

Table 3. Design of full-definition standards and idea-unit standards

Full-definition standards	Idea-unit standards
Claim: An assertion or conclusion explaining the investigated phenomenon	[a] (1) Write a complete sentence (2) Provide a causal account of the investigated phenomenon (e.g. how the independent variable affects the dependent variable)
Evidence: Scientific data supporting the claim. Data often come from observations or experiments. More than one piece of evidence may be needed	[a] (1) Evidence often comes from observations or experiments (2) Evidence must support the claim
Reasoning: A justification that links the claim and evidence and shows why the data qualify as evidence to support the claim using appropriate and sufficient scientific principles	[a] (1) Use adequate scientific principles (2) The use of concepts or principles needs to connect the claim and the evidence

[a]Same as the full-definition standards.

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Data Collection

Two instruments were used to examine the students' progress in terms of two learning outcomes: (1) a biology concept test assessing students' understanding of conceptions of related biology topics was administered before and after 10 weeks of instruction, and (2) the students' investigation sheets for each unit were collected and rated to assess the quality of their explanations.

To answer the first research question regarding whether adding specific cognitive or metacognitive scaffolds may benefit conceptual understanding, a biology concept test was developed. Adequately designed inquiry activities and effective scaffolds may help students link scientific concepts with the phenomenon under investigation rather than inferring based on common sense evidence or reasoning that may result in alternative conceptions (Talanquer, 2006). Thus, adequate conceptual understanding was considered as being able to select and differentiate scientific conceptions from alternative ones.

The item development procedure involved rigorous qualitative alignments (DeBoer, Herrmann-Abell, Wertheim, & Roseman, 2009) that included the following steps: (1) Unpacking key conceptions for explaining each phenomenon under study (see Table 2) while developing course instruction, topics for investigation, and cognitive question prompts. These key conceptions were then used for item development. (2) Identifying naïve conceptions and alternative ideas that students have and use to develop distracters in test items. (3) Six seventh-grade students who did not participate in this study were invited for interviews in order to receive feedback for determining item readability and appropriateness. Once a draft of the items was formed, (4) the items were evaluated and revised to ensure the key conceptions specified in step (1) were both necessary and sufficient to correctly answer the items (DeBoer et al., 2008). The same panel members who developed the course instructions and cognitive scaffolds worked collaboratively throughout the entire item development procedure. The final version of the biology concept test consisted of 23 multiple-choice questions requiring students to differentiate and select the correct scientific conceptions from distracters that presented common alternative conceptions. An example question for unit 3 is illustrated as follows:

What is the main driving force of transpiration?

A. Water absorption at the roots creates pressure to push water upward which then escapes from the leaves.
B. Evaporation of water from stomas of leaves brings up water in vascular bundles.
C. Consumption of water by photosynthesis at the leaves creates a negative pressure that moves water upwards in vascular bundles.
D. Respiration takes place and drives the movement of water.

Distracters for the items were developed based on common alternative ideas observed in two participating teachers' classrooms. One common source of alternative conceptions was that students use inadequate concepts for explanation (Talanquer,

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

2006). For example, some students may have been confused about the relationships among plant respiration, photosynthesis, and transpiration when those terms were introduced in one unit. Thus, there might be some students who mistakenly used inadequate conceptions, such as respiration, to explain the driving force of transpiration (distracter D).

Items of the biology concept test cover the five topics of inquiry activities implemented in this study. Depending on the number of key conceptions in a topic, 4 items were developed for units 2 and 5, while units 1, 3, and 4 each contained 5 items. The total possible score on the test was 23 points, and the pretest and posttest were identical. Students who received a higher score on the biology concept test were considered to possess more adequate conceptual understanding resulting from the instruction. The Cronbach alphas for the pretest and posttest were 0.68 and 0.71, respectively. The slightly low Cronbach alpha for the pretest may be attributed to using a limited number of items to cover five biology topics. A one-way analysis of variance (ANOVA) performed on the pretest scores showed no significant differences among the three groups ($F(2, 170) = 2.09$, $p = .127$). Thus, the students in the three instructional groups had the same entry level of content knowledge.

To gather reliable data for comparisons across the three conditions, the base rubric used in McNeill et al.'s study (2006) was adopted. A rubric scoring system with the criteria for the scoring levels of the explanation components was also developed for each unit by following McNeill et al.'s (2006) suggestions about developing specific explanation rubrics. The rating quality for the components of the explanation included both conceptual adequacy and the appropriate explanation structure to support the claim. Each component was rated as complete (2 points), partially complete (1 point), or incomplete (0 point), and examples of students' work were added. A specific explanation rubric for unit 2 is presented in Appendix 2 as an example. The panel members reached a consensus on the rubrics by collectively scoring the investigation sheets of a class and selecting examples of complete, partially complete, and incomplete statements from the students' investigation sheets. Then, all raters scored the remaining investigation sheets independently using the rubric. The graduate student from the panel sampled and rerated 30% of the rated artifacts to establish inter-rater reliability. The explanations of a class were rerated and discussed if the inter-rater reliability was below 90%. This rating process was repeated for each unit, with final inter-rater reliability exceeding 90% for all classes and units.

To explore the effect of metacognitive evaluation instruction on resolving students' inadequate evaluation, the deviation between peer-rating and researcher-rating for the M + E group was assessed by subtracting the researcher-rating from the peer-rating of a specific students' explanation. A positive value indicates an overestimation, whereas a negative value represents an underestimation. Finding a near zero deviation indicates that students could accurately evaluate their peers' work, while a greater deviation suggests poorer understanding of the criteria.

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Results

Effects of Scaffolding Designs on Students' Growth of Content Knowledge

Table 4 presents the means (M) and standard deviations (SD) of the biology content pretest and posttest scores. The pair-sampled t-tests showed that students in all groups exhibited a significant learning gain from the pretest to the posttest. A one-way analysis of covariance was conducted using the posttest as a dependent variable and the pretest as a covariate. The analysis showed a significant main effect for instructional mode ($F(2, 156) = 5.26, p < .01$, partial $\eta^2 = .063$). Pairwise comparisons of the adjusted means revealed that the C + E group ($M = 15.59$) outperformed the other two groups. No significant difference was found between the M + E ($M =$ 14.20) and E-only ($M = 14.09$) groups.

Influence of Scaffolding Design on Students' Explanations

Table 5 summarizes the descriptive statistics for the average scores of the five units for claims, evidence, and reasoning. The average scores for claims were higher for the M + E group than for the C + E and E-only groups. The students who received cognitive prompts (C + E) outperformed the students in the M + E and E-only groups on both evidence and reasoning. The mean difference in the evidence score was minor between the M + E and E-only groups; however, the M + E students showed better reasoning scores than the E-only students.

To test whether there were differences in the students' explanations for the different instructional modes during the units, separate repeated measures analyses of variance (RMANOVAs) were conducted on the claim, evidence, and reasoning scores. Each RMANOVA was 3 × 5; instructional mode (C + E, M + E, and E-only) was the between-subjects factor, and time was the within-subjects factor. The results for the effects of instructional mode, time, and the interaction between these factors on claims, evidence, and reasoning are summarized in Table 6. According to Table 6, for the claim scores, there was a significant main effect for instructional mode but not for time. Students' evidence scores showed both significant main effects for instructional mode and time. Similar results were found for the reasoning scores, for which there were significant main effects for instructional mode and time. In addition, significant interaction effects of instructional mode × time were observed

Table 4. Descriptive statistics of pretest and posttest content knowledge

	Pre		Post		M		
	n	M (SD)	n	M (SD)	Difference	t	p
Instructional mode							
E-only	59	11.39 (4.22)	55	14.29 (3.84)	2.90	6.98	<.001
C + E	55	10.22 (3.76)	55	15.34 (3.23)	5.12	11.85	<.001
M + E	59	10.10 (3.25)	58	13.78 (3.54)	3.68	9.50	<.001

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Table 5. Descriptive statistics for the average scores for claims, evidence, and reasoning

	Claim[a]		Evidence[a]		Reasoning[a]	
	n	*M* (SD)	*n*	*M* (SD)	*n*	*M* (SD)
Instructional mode						
E-only	57	1.32 (0.28)	57	0.96 (0.32)	56	0.32 (0.34)
C + E	49	1.29 (0.24)	49	1.17 (0.27)	49	0.71 (0.31)
M + E	57	1.45 (0.31)	57	0.85 (0.39)	57	0.50 (0.39)

[a]Maximum score = 2.0.

Table 6. Effects of instructional mode and time on claims, evidence, and reasoning

	Effect	*F*	Partial η^2
Claim	Instructional mode	5.13**	.06
	Time	1.44	.009
	Instructional mode × time	7.45***	.09
Evidence	Instructional mode	7.04***	.14
	Time	43.81***	.22
	Instructional mode × time	4.77***	.06
Reasoning	Instructional mode	16.67***	.17
	Time	26.55***	.14
	Instructional mode × time	2.38*	.03

*$p < .05$.
**$p < .01$.
***$p < .001$.

for claims, evidence, and reasoning. These interaction effects are discussed in more detail below.

Effects of Scaffolding Designs on Explanations Over Units

To explore whether the scaffolds that the students received during the unit influenced their explanations, Figures 1–3 chart the mean scores for claims, evidence, and reasoning, respectively, for the three instructional modes based on the investigation sheets the students completed in the five units. The simple main effect of time was examined separately for each group and is reported in the following sections to reveal whether the quality of the students' explanations changed over time. The mean differences for each unit were tested for the three instructional modes, as summarized in Appendix 3.

An analysis of the students' claim scores indicated a significant trend in the M + E group ($F(4, 224) = 14.91$, $p < .001$, partial $\eta^2 = .21$) over units but not in the C + E ($F(4, 192) = 1.64$, $p = .17$, partial $\eta^2 = .03$) and E-only ($F(4, 224) = 2.29$, $p = .06$, partial $\eta^2 = .04$) groups. As shown in Figure 1, the claim scores of the M + E

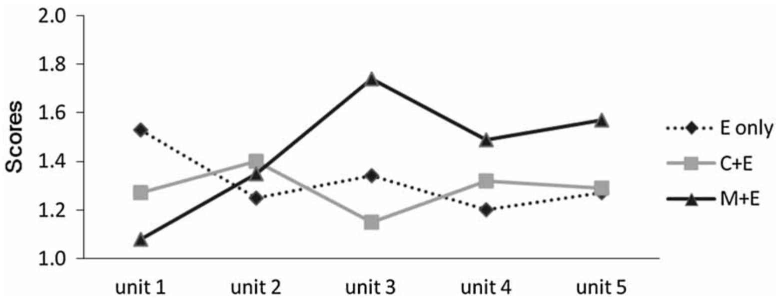

Figure 1. Claim scores over the units

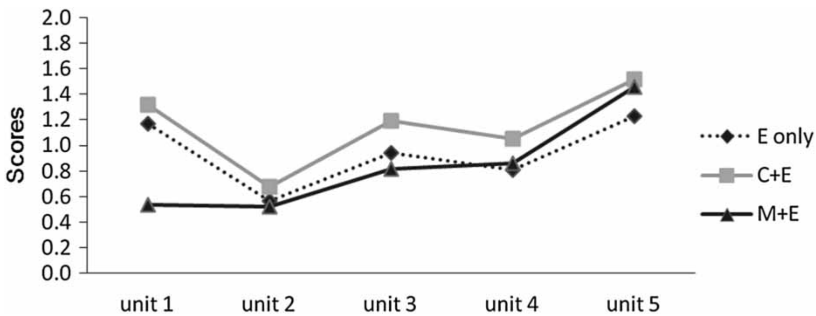

Figure 2. Evidence scores over the units

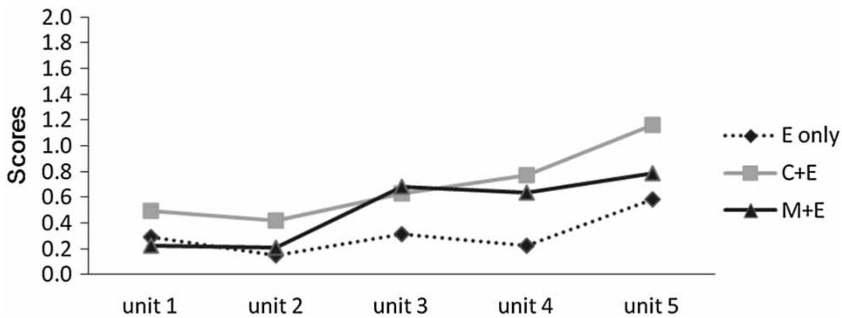

Figure 3. Reasoning scores over the units

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

group increased continuously from units 1 through 3 and remained high for units 4 and 5. Analyses of the mean differences for each unit (see Appendix 3) revealed that the M + E group initially and significantly underperformed the other two groups (partial η^2 = .12) in unit 1. By unit 2, the M + E group students were as proficient as their counterparts (partial η^2 = .007) and outperformed them throughout the remaining units (unit 3: partial η^2 = .16; unit 4: partial η^2 = .06; unit 5: partial η^2 = .05).

A significant improvement trend was observed for all groups in their evidence scores (E-only: $F(4, 224) = 12.37$, $p < .001$, partial η^2 = .18; C + E: $F(4, 192) = 18.32$,

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

$p < .001$, partial $\eta^2 = .28$; M + E: $F(4, 224) = 22.70$, $p < .001$, partial $\eta^2 = .29$). As shown in Figure 2, both the C + E and E-only groups started with higher evidence scores than the M + E group in unit 1 (partial $\eta^2 = .17$) but then experienced a drop in unit 2. Whereas the evidence scores of all groups showed increases from units 2 to 3 and from units 4 to 5, analyses of the mean difference for each unit (see Appendix 3) showed that the C + E group outperformed the E-only group in units 3, 4, and 5 (partial $\eta^2 = .06$, .04, and .04, respectively). The M + E group underperformed the C + E group in units 1 and 3; however, the gap was reduced and the difference became non-significant in units 4 and 5. In comparison with the E-only group, the differences between the evidence scores of the M + E and E-only groups were non-significant in units 2, 3, and 4, but the former group significantly outperformed the E-only group in unit 5.

All instructional modes showed a significant developing trend over units for the reasoning scores (E-only: $F(4, 220) = 5.97$, $p < .001$, partial $\eta^2 = .10$; C + E: $F(4, 192) = 12.28$, $p < .001$, partial $\eta^2 = .20$; M + E: $F(4, 224) = 11.00$, $p < .001$, partial $\eta^2 = .16$). As shown in Figure 3, the reasoning scores of the C + E group increased continuously from units 2 through 5. For the E-only group, the increase did not appear until unit 5. The reasoning scores of the M + E group showed a significant increase in unit 3 but leveled off in units 4 and 5. Analyses of the mean difference for each unit (see Appendix 3) indicated that the C + E group outperformed the other groups in units 1 and 2 (partial $\eta^2 = .04$ and .05, respectively). The gap between the C + E and M + E groups was reduced and the two groups outperformed the E-only group (unit 3: partial $\eta^2 = .07$; unit 4: partial $\eta^2 = .10$) in units 3 and 4. In unit 5, the students in the C + E group outperformed their counterparts (partial $\eta^2 = .09$), and the difference between the M + E and E-only groups was non-significant.

To explore the effect of metacognitive evaluation instruction on improving students' understanding of quality explanations, the averaged scores for difference of peer/researcher-rating for claims, evidence, and reasoning are charted in Figure 4. Separated single-sample *t*-tests were also conducted for each unit for claims, evidence, and reasoning to examine if the difference in peer/researcher-rating was greater than zero. Means and standard deviations for the differences and the results of the single-sample *t*-tests are summarized in Appendix 4. Separated RMANOVAs were also

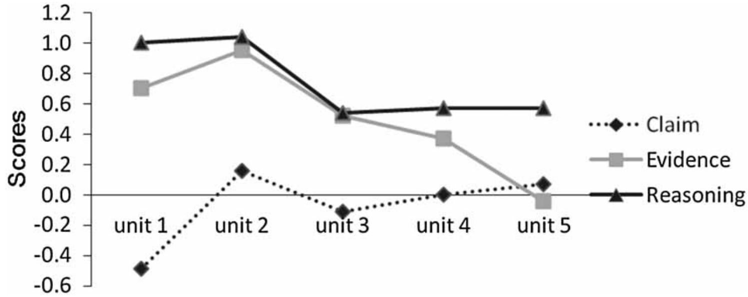

Figure 4. Differences in peer/researcher-rating for claims, evidence, and reasoning over units

performed on the differences in peer/teacher-rating for claims, evidence, and reasoning using time as the within-subjects factor.

Overall, the trends in Figure 4 indicate that the students underestimated the quality of claims but overestimated when rating their peer's quality of evidence and reasoning. The results of the RMANOVAs showed a significant effect of time on reducing the underestimation of claims ($F(4, 224) = 7.96$, $p < .001$, partial $\eta^2 = .12$) and the overestimation of evidence ($F(4, 224) = 13.67$, $p < .001$, partial $\eta^2 = .20$) as well as reasoning ($F(4, 220) = 6.61$, $p < .001$, partial $\eta^2 = .11$). As seen in Figure 4, the difference in peer/researcher-rating for claims was significantly underestimated in unit 1 ($M = -0.45$, $t = -5.31$, $p < .001$) but reached statistical insignificance in units 2 to 5. As for evidence, the overestimation was gradually decreased after unit 2 and reached statistical insignificance with zero in unit 5 ($t = -0.93$, $p >$ .05). The overestimation of reasoning decreased from units 2 to 3 but did not reduce further in the subsequent units.

In addition, analyses of Pearson's product-moment correlations coefficients were carried out to examine if the differences in peer/researcher-rating was associated with the M + E students' quality of explanation; the results are summarized in Appendix 5. Moderate-to-strong negative correlations were observed for claims, evidence, and reasoning in all units, except for the claim score for unit 1.

Discussion

Explicit Cognitive Prompts Better Facilitate Knowledge Integration and Result in Greater Growth of Content Knowledge

The first research question concerns whether the three scaffolding designs (C + E, M + E, and E-only) have different effects on students' growth of biology concepts. These findings suggest that cognitive prompt questions, adapted specifically for each topic, successfully demonstrate to students the connections between concepts and explanation construction. Responding to these questions helped direct the students' attention toward salient concepts for that topic and activated related scientific principles before constructing explanations. In addition, the hints embedded in reasoning made the association between the content knowledge and the scientific explanation explicit by reminding learners to consider their responses to the prompt questions while forming their reasoning. Incorporating cognitive prompts with the explanation scaffolds facilitated knowledge integration, in which the students connected the concepts learned in the classroom with information synthesized from the inquiry activity and, in turn, resulted in better learning outcomes of content knowledge in the posttest. The effect of cognitive prompts on facilitating knowledge integration was also supported because the C + E students demonstrated better reasoning quality through their use of appropriate concepts to support their claims, as discussed in the next sections. The positive effect of cognitive prompts on facilitating conceptual understanding echoes the findings of Bulu and Pedersen (2010), in which learning was supported in a problem-solving situation. Similar to the function

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

of Chin and Osborne's (2010) scaffolding design in terms of having learners generate cognitive questions, the content-specific scaffolds used in this study have shown their effectiveness in resolving cognitive obstacles and facilitating learning of science content.

For the metacognitive evaluation instruction, the original hypothesis was that the scaffolding design may improve the quality of explanation while also improving conceptual understanding by enhancing knowledge integration during explanation construction. Significant improvement was observed from the pretest to posttest; nonetheless, the metacognitive approach did not show any additional effect on facilitating conceptual understanding in comparison with the explanation scaffolds alone. For the metacognitive evaluation instruction, knowledge integration was facilitated implicitly by using idea-unit standards and domain-specific guidance to remind the students of the criteria for explanations and by discussing how to revise partially complete and incomplete examples based on the criteria. Thus, the implicit facilitation may only benefit those students who possess and recognize which content knowledge should be used for explanation construction and only have a limited effect on enhancing knowledge integration for students who experience functional fixedness or functional reduction. Unlike the effectiveness of cognitive prompts in explicitly revealing the association between content knowledge and explanation construction, the indirect supports implicitly embedded in the metacognitive evaluation instruction did not show an added effect of enhancing conceptual understanding. The observation about lacking an effect of metacognitive evaluation on enhancing conceptual understanding has not been reported in related studies (e.g. Chang et al., 2010; Davis, 2000) and will require further investigation.

Cognitive and Metacognitive Scaffolding Designs Revealed Different Effects on Students' Quality of Claims, Evidence, and Reasoning

The second research question involved whether the scaffolding designs had different effects on students' quality of explanation over units.

Effects of the cognitive prompt scaffolding: The learning trajectories of the E-only group revealed that the explanation scaffolding improved the students' quality of evidence but not that of claims and reasoning. By adding the cognitive prompts, quality of evidence and reasoning was significantly improved. The difference in performance of the C + E and E-only students may result from the ability of cognitive prompts to address their procedural difficulties of cognitive reasoning related to functional fixedness and functional reduction.

Qualitative analyses of the students' investigation sheets showed that a relatively higher ratio of E-only students showed signs of functional reduction while constructing evidence and reasoning. Taking the transpiration topic as an example, some of them used only common experience as evidence (e.g. during winter time, trees lose leaves to cut down water consumption) without referring to the first-hand data collected during their investigation. Others reduced the task complexity by providing vague descriptions as evidence, such as 'no leaves, no water

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

consumption', or offered insufficient evidence, such as 'the decreased water level was greater in setting A than in setting B' without referring to the difference in the number of leaves in the two settings. Similar findings were observed for reasoning where the E-only students repeated evidence as reasoning without referring to scientific principles, or repeated their claim by stating 'the number of leaves would affect the speed of transpiration' without connecting the claim to the evidence. These difficulties encountered by the E-only group were similar to the obstacles reported in previous findings (e.g. McNeill et al., 2006; Sampson et al., 2011; Sandoval & Reiser, 2004). By incorporating cognitive prompts, a higher ratio of students in the C + E group provided more evidence by referring to their observation data that depicted relations between the number of leaves and the decreased water level in the experimental and control settings. In terms of reasoning, more C + E students were capable of using underpinning concepts (e.g. driving forces of transpiration) to justify how their observations indicated the effect of transpiration. These findings show that the cognitive prompts in the forms of questions and hints were effective in helping the students recall or recognize the salient features and concepts of a specific topic (Bulu & Pedersen, 2010), which might otherwise be missed by some students due to function reduction. In addition, combining cognitive prompts with the explanation scaffolding was effective in making explicit the associations between the underpinning concepts and the information generated from the inquiry activity and, therefore, facilitating inference generation (Chin & Osborne, 2010). This explicit facilitation helps to overcome functional fixedness and functional reduction by redirecting students' attention away from common sense reasoning and guiding them to select and apply appropriate evidence and scientific principles to support their claims, simultaneously enhancing knowledge integration.

Regarding claims, the explanation scaffolding helped the students of both groups obtain partial scores for claims, but the quality of their claims did not further improve when the cognitive prompt scaffolding was added. The insignificant difference between the C + E and E-only groups may be due to the prompt questions and hints not addressing what constitutes an adequate claim to explain the investigated phenomenon. Discussing the full definition of claims with complete examples may not provide sufficient information to help students understand what needs to be included to qualify as an 'accurate and complete' claim. Thus, using the full-definition standards and discussing with complete examples revealed only minimal effects on helping the students distinguish partially complete from complete claims, and offered limited cues for students to identify which part of their claims needed to be improved to receive full credits. The findings of the present study indicate that even when students have received instruction on the related concepts, additional cognitive scaffolding is needed. A positive effect of cognitive prompts on enhancing quality of evidence and reasoning was observed and is congruent with findings of previous studies (e.g. Bulu & Pedersen, 2010; Chin & Osborne, 2010; Sandoval, 2003); however, the effect differs for improving the quality of claims, which was not reported in the related studies.

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Effects of the metacognitive evaluation instruction: The present results regarding the M + E students indicate that the degree to which learners better evaluate the quality of claims, evidence, and reasoning and improve their scientific explanations can be enhanced using a metacognitive evaluation approach. Highlighting key features or components of the evaluation criteria and providing opportunities to analyze strong and weak examples using the refined criteria helped the students better interpret the meaning of 'accurate and complete' explanations. This finding is in line with the results of Davis's (2000) and Chang et al.'s (2010) studies. When further testing the effect of the idea-unit standards (Lipko et al., 2009), the present results revealed that the students became better at recognizing inconsistencies between their peers' explanations and the corresponding criteria during peer-evaluation. The supporting evidence includes the gradually closed gaps in the differences between the peer/researcher-ratings for claims, evidence, and reasoning over units. Along with the cyclical process of improving evaluation accuracy using the idea-unit standards, the students also gradually refined their personal criteria for self-monitoring and were more capable of identifying which parts of their explanations must be improved in order to plan the next cycle of explanation construction (Winne & Hadwin, 1998). The influence of the cyclical self-regulatory process is supported by the strong associations between the students' improved understanding of scientific explanation, as demonstrated by accurate peer-evaluation, and their actual performance on constructing scientific explanations.

Although the metacognitive evaluation instruction was effective in helping the students internalize the refined criteria and improve their understanding of the general guidelines for effective scientific explanations, applying and transferring the general guidelines to other topic-specific contexts through a self-regulatory process may require multiple exercises. The trajectory for reaching accurate evaluation varied among claims, evidence, and reasoning as well. For instance, the underestimation of claims was reduced quickly after one cycle; however, reducing the overestimation of evidence and reasoning required more units to close the gap. The students may have vague perceptions of criteria of evidence and reasoning at the beginning of the instruction because these criteria are rarely explicit in the science classroom. This vague understanding of the criteria of evidence and reasoning might require multiple learning cycles to internalize adequate criteria and improve performance. This need for multiple learning cycles may also explain why a significant difference in the evidence scores of the M + E and E-only groups was not observed until unit 5. Learning with domain-general scaffolds such as the metacognitive evaluation instruction results in a gradual learning trajectory in comparison to the cognitive prompt scaffolding, which was adapted to topic-specific features to provide explicit facilitation for knowledge integration.

The present study adopted a finer-grained lens to examine whether and how the designs of cognitive and metacognitive scaffolding influence learning of explanation and science content. This study has reported the different effects of cognitive and metacognitive scaffolding on enhancing the quality of claims, evidence, and reasoning. The variations in trends for the changing process of explanation quality and the gradients of the closing gaps between the peer/researcher-ratings over units have not been reported elsewhere. These results will contribute to current understanding of the learning trajectory of scientific explanation.

Conclusions and Implications

The construction of scientific explanation is a complex task for middle school students. The present study teases out students' difficulties stemming from their lack of understanding of scientific explanation, flawed cognitive reasoning, or poor evaluation skills, and presents important findings related to the design of scaffolds to target these challenges in constructing adequate scientific explanations. The findings reveal that explicitly introducing the rationale and essence of scientific explanation and supplementing written explanation scaffolds may enhance learners' understanding of the role of explanations in scientific inquiry to some extent; however, during complex tasks such as constructing scientific explanations, adding cognitive prompts specifically adapted to each topic has the added effect of overcoming cognitive challenges and enhancing knowledge integration. However, the facilitation of the current design of cognitive prompts has selective effects: it improved the quality of evidence and reasoning but had limited effects on claims. Science teachers who aim to enhance the growth of content knowledge or improve students' ability of using evidence and synthesizing reasoning can integrate topic-specific cognitive prompts with the explanation scaffolding.

Learners' inadequate evaluation was addressed through a metacognitive evaluation approach. The findings of this study indicate that introducing full definitions of scientific explanations may not be sufficient for those students with inadequate internal standards. To address inadequate evaluation, highlighting the essential features required to receive full credits and offering multiple opportunities to identify inconsistencies between the criteria of evaluation and the to-be-evaluated works may help the students refine their internal standards. Improving evaluation accuracy and helping students identify the aspects of their work that need to be improved in order to plan the next learning cycle may have a long-term benefit for their learning. The present study used cyclical metacognitive evaluation as an alternative approach to enhancing scientific explanations through refining students' internal standards and prompting monitoring and evaluation. Compared with other scaffolding designs that directly provide metacognitive knowledge or training of skills, the metacognitive intervention in this study is rather implicit and may not be the most effective way of engaging metacognitive thinking. Future studies should compare the effectiveness of this metacognitive approach in terms of improving performance and self-regulated learning with other scaffolding designs such as introducing metacognitive strategic knowledge (Zohar & Peled, 2008) or prompting metacognitive skills such as planning, monitoring, and evaluation (Chen, 2010). Also, more research is needed to understand how middle school students with different levels of metacognitive skills can be supported by this metacognitive evaluation instruction.

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

In this study, we do not discuss whether the effects of these scaffolding designs varied in different content areas; however, the amount of content knowledge possessed by students in a particular topic (e.g. McNeill et al., 2006) or the use of data (e.g. first-hand versus second-hand) may influence their learning of claims, evidence, or reasoning. Future studies may investigate the effects of domain-specific versus general scaffolding on facilitating the learning of content knowledge and scientific explanations across topics, disciplines, different levels of task complexity, or contexts of scientific investigations (e.g. simulated versus hands-on experiments). When assessing students' gains in conceptual understanding across diverse topics using a multiple-choice, knowledge assessment, achieving an acceptable reliability may become challenging. Future researchers can consider adding qualitative data (e.g. collecting open-ended responses, Chang et al., 2010; McNeill et al., 2006) to triangulate data of multiple-choice items in order to improve the quality of the evidence.

The support provided by both the cognitive and metacognitive scaffoldings is substantial, and one cannot be replaced by the other. Combining both cognitive and metacognitive scaffolds might be more effective for improving scientific explanations by not only directing cognitive attention but also refining the students' understanding of the criteria of explanations and enhancing self-monitoring. However, incorporating both types of scaffolds might create cognitive overload and require a longer time for students to achieve proficiency. Future studies should test the combinational effect of cognitive and metacognitive scaffoldings and their interactions with cognitive load on performance. The learning trajectories in Figures 1–3 indicate that all instructional modes require multiple sessions for seventh-grade students to adapt to the instruction. However, whether the learning trajectory is saturated by the fifth unit of instruction is still unknown. More exercises might be needed because inquiry is complex and certain aspects (such as reasoning) may be more difficult than others. The present study aimed to individually investigate the effects of cognitive and metacognitive scaffolding using continuous supports. Future studies should pay careful attention to the relation between the duration of scaffolds and the timing of fading the scaffolding as well as their effects on learning. The optimal conditions for the duration of support and the timing of its removal might differ for cognitive, metacognitive, and combined scaffolds.

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Funding

Funding for this research work was supported by the National Science Council, Taiwan, under grant numbers [NSC 99-2511-S-009-002-MY2 and NSC 101-2511-S-009-001].

References

Berland, L. K., & Reiser, B. J. (2009). Making sense of argumentation and explanation. *Science Education*, *93*, 26–55.

Berthold, K., Eysink, T. H. S., & Renkl, A. (2009). Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instructional Science*, *37*, 345–363.

Braaten, M., & Windschitl, M. (2011). Working toward a stronger conceptualization of scientific explanation for science education. *Science Education*, *95*, 639–669.

Bulu, S. T., & Pedersen, S. (2010). Scaffolding middle school students' content knowledge and ill-structured problem solving in a problem-based hypermedia learning environment. *Educational Technology Research and Development*, *58*, 507–529.

Chang, H.-Y., Quintana, C., & Krajcik, J. S. (2010). The impact of designing and evaluating molecular animations on how well middle school students understand the particulate nature of matter. *Science Education*, *94*, 73–94.

Chen, C.-H. (2010). Promoting college students' knowledge acquisition and ill-structured problem solving: Web-based integration and procedure prompts. *Computers & Education*, *55*, 292–303.

Chin, C., & Osborne, J. (2010). Students' questions and discursive interaction: Their impact on argumentation during collaborative group discussions in science. *Journal of Research in Science Teaching*, *47*, 883–908.

Chiu, J., & Linn, M. (2012). The role of self-monitoring in learning chemistry with dynamic visualizations. In A. Zohar & Y. J. Dori (Eds.), *Metacognition in science education: Trends in current research* (pp. 133–163). Dordrecht: Springer.

College Board. (2009). *Science: College board standards for college success*. New York, NY: Author.

Davis, E. A. (2000). Scaffolding students' knowledge integration: Prompts for reflection in KIE. *International Journal of Science Education*, *22*, 819–837.

DeBoer, G. E., Herrmann-Abell, C. F., Gogos, A., Michiels, A., Regan, T., & Wilson, P. (2008). Assessment linked to science learning goals: Probing student thinking through assessment. In J. Coffey, R. Douglas, & C. Stearns (Eds.), *Assessing student learning: Perspectives from research and practice* (pp. 231–252). Arlington, VA: NSTA Press.

DeBoer, G. E., Herrmann-Abell, C. F., Wertheim, J., & Roseman, J. E. (2009). *Assessment linked to middle school science learning goals: A report on field test results for four middle school science topics*. Paper presented at the national association for research in science teaching annual conference, Garden Grove, CA.

Ford, C. L., & Yore, L. D. (2012). Toward convergence of critical thinking, metacognition, and reflection: Illustrations from natural and social sciences, teacher education, and classroom practice. In A. Zohar & Y. J. Dori (Eds.), *Metacognition in science education: Trends in current research* (pp. 251–271). Dordrecht: Springer.

Furió, C., Calatayud, M. L., Bárcenas, S. L., & Padilla, O. M. (2000). Functional fixedness and functional reduction as common sense reasonings in chemical equilibrium and in geometry and polarity of molecules. *Science Education*, *84*, 545–565.

Lipko, A. R., Dunlosky, J., Hartwig, M. K., Rawson, K. A., Swan, K., & Cook, D. (2009). Using standards to improve middle school students' accuracy at evaluating the quality of their recall. *Journal of Experimental Education: Applied*, *15*, 307–318.

McNeill, K. L., & Krajcik, J. (2008). Scientific explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning. *Journal of Research in Science Teaching*, *45*, 53–78.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences*, *15*, 153–191.

National Research Council. (2011). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.

OECD. (2013). *PISA 2015 draft science framework*. Paris: Author.

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Osborne, J. F., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, *41*, 994–1020.
Osborne, J. F., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? *Science Education*, *95*, 627–638.
Rawson, K. A., & Dunlosky, J. (2007). Improving students' self-evaluation of learning for key concepts in textbook materials. *European Journal of Cognitive Psychology*, *19*, 559–579.
Sampson, V., & Clark, D. B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education*, *92*, 447–472.
Sampson, V., Grooms, J., & 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. *Science Education*, *95*, 217–257.
Sandoval, W. (2003). Conceptual and epistemic aspects of students' scientific explanations. *The Journal of the Learning Sciences*, *12*, 5–51.
Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction*, *23*, 23–55.
Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, *88*, 345–372.
Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology*, *99*, 285–296.
Talanquer, V. (2006). Commonsense chemistry: A model for understanding students' alternative conceptions. *Journal of Chemical Education*, *83*, 811–816.
Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ: Lawrence Erlbaum.
Zohar, A., & Peled, B. (2008). The effects of explicit teaching of metastrategic knowledge on low- and high-achieving students. *Learning and Instruction*, *18*, 337–353.

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Appendix 1. An example of explanation scaffolding for students' investigation sheets

E-only	C + E	M + E
	Prior to forming your explanation, think about the following prompt questions: (1) How does water leave a plant during transpiration? (What is the evidence?) (2) How does the water level of the experimental group differ from that of the control group? What drives plant transpiration?	
Topic: How does the number of leaves affect the speed of transpiration? Explain your reasons	[a]	[a]
Claim: Write a sentence describing how the number of leaves affects the speed of transpiration	[a]	[a] Peer-evaluation of quality of scientific explanation ☐ Complete ☐ Partially complete ☐ Incomplete Do you agree/disagree with this rating? (circle one)
Evidence: Provide evidence (data that come from observations or experiments) supporting your claim for how the number of leaves affects the speed of transpiration Evidence 1: Evidence 2:	[a]	[a] Peer-evaluation of quality of scientific explanation ☐ Complete ☐ Partial complete ☐ Incomplete Do you agree/disagree with this rating? (circle one)
Reasoning: Use concepts you learned previously to describe how the evidence supports your claim for how the number of leaves affects the speed of transpiration	[a] Note: *Use your responses to the prompt questions to help you synthesize your reasoning*	[a] Peer-evaluation of quality of scientific explanation ☐ Complete ☐ Partial complete ☐ Incomplete Do you agree/disagree with this rating? (circle one)

[a]The structure and content are the same as the E-only group.

Appendix 2. Specific explanation rubric for unit 2

	Level		
Component	0	1	2
Claim	Does not make a claim or makes an inaccurate claim	Makes an accurate but incomplete claim (e.g. implies but does not specifically describe how temperature affects the activity of amylase)	Makes an accurate and complete claim that provides a causal account for the investigated phenomenon
Evidence	Does not provide evidence or only provides inappropriate evidence	Provides at least one piece of appropriate evidence; may include some inappropriate evidence	Provides three pieces of appropriate evidence (e.g. describes data for activity of amylase under 5°C, 40°C, and 90°C) and contains no inappropriate evidence
Reasoning	Does not provide reasoning; provides reasoning that does not link evidence to claim (e.g. repeats the evidence)	Includes some scientific principles but not sufficient or includes appropriate and sufficient principles but only links to either claim or evidence	Includes appropriate and sufficient scientific principles (e.g. indicates the role of amylase and the process of amylolysis) to provide reasoning that links evidence to the claim

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015

Appendix 3. Pairwise comparisons for mean differences among the three instructional modes

	Claim			Evidence			Reasoning		
	F	Pairwise	partial η^2	*F*	Pairwise	partial η^2	*F*	Pairwise	partial η^2
Unit 1	12.60***	E-only > C + E, E-only > M + E, C + E > M + E	.12	18.68***	E-only > M + E, C + E > M + E	.17	3.66*	C + E > M + E	.04
Unit 2	0.65		.007	1.86		.02	4.59*	C + E > E-only, C + E > M + E	.05
Unit 3	15.89***	M + E > E-only, M + E > C + E	.16	5.27**	C + E > E-only, C + E > M + E	.06	6.18**	C + E > E-only, M + E > E-only	.07
Unit 4	4.92**	M + E > E-only	.05	3.18*	C + E > E-only	.04	9.85***	C + E > E-only, M + E > E-only	.10
Unit 5	5.12**	M + E > E-only M + E > C + E	.05	3.62*	C + E > E-only, M + E > E-only	.04	9.08***	C + E > E-only, C + E > M + E	.09

*$p < .05$.

**$p < .01$.

***$p < .001$.

Appendix 4. Descriptive statistics and summary of the single-sample *t*-tests for difference of peer/researcher-rating for the M + E group

	Unit 1		Unit 2		Unit 3		Unit 4		Unit 5	
	M (SD)	*t*	*M* (SD)	*t*	*M* (SD)	*t*	*M* (SD)	*t*	*M* (SD)	*t*
Claim	−0.45 (0.67)	−5.31***	0.19 (0.80)	1.89	−0.11 (0.66)	−1.36	0.02 (0.73)	0.18	0.06 (0.74)	0.68
Evidence	0.73 (0.78)	7.31***	0.94 (0.78)	9.57***	0.51 (0.88)	4.55***	0.39 (0.75)	3.97***	−0.08 (0.68)	−0.93
Reasoning	1.06 (0.81)	10.39***	1.05 (0.75)	11.09***	0.48 (0.76)	5.00***	0.58 (0.91)	4.85***	0.56 (0.69)	6.42***

***$p < .001$.

Appendix 5. Relationship between difference of peer/researcher-rating and quality of explanation for the M + E group

	Unit 1 ($n = 62$)	Unit 2 ($n = 63$)	Unit 3 ($n = 62$)	Unit 4 ($n = 59$)	Unit 5 ($n = 63$)
Claim	$-.20$	$-.62^{***}$	$-.63^{***}$	$-.49^{***}$	$-.68^{***}$
Evidence	$-.57^{***}$	$-.42^{***}$	$-.57^{***}$	$-.50^{***}$	$-.61^{***}$
Reasoning	$-.65^{***}$	$-.27^{*}$	$-.52^{***}$	$-.71^{***}$	$-.54^{***}$

$^{*}p < .05$.

$^{***}p < .001$.

Downloaded by [Flinders University of South Australia] at 00:30 28 January 2015