



EDUCATIONAL RESEARCH

Assessment of self-explaining effect in a large enrolment General Chemistry course



Adrian Villalta-Cerdas¹, Santiago Sandi-Urena^{*,2}

Department of Chemistry, University of South Florida, CHE 205, Tampa, FL 33620, USA

Received 10 July 2015; accepted 8 October 2015

Available online 5 February 2016

KEYWORDS

Self-explaining;
Classroom study;
Chemistry education;
Scientific
explanation;
General Chemistry

Abstract Self-explaining refers to the generation of inferences about causal connections between objects and events for one's own consumption. Self-explaining is amongst the practices of science deemed essential for scientific competence; therefore, a valued learning outcome in itself. Nonetheless, generation of authentic explanations is seldom promoted in college science instruction. This work examined the effect of engagement in self-explaining on conceptual understanding of chemistry. Learning and performance tasks were completed individually in the classroom ecology of a large-enrolment General Chemistry course in the US. The study spanned a period of five semesters including pilot-tests and replications. The self-explaining intervention followed a multi-condition comparison design that used performance on a post-test to assess learning. Students were randomly assigned to the following conditions: reviewing a correct explanation, explaining correct or incorrect answers, explaining agreement with answers produced by others, and explaining their own answers. A cohort of students who underwent standard instruction with no intervention and had prepared for formal examination served as reference. The self-explaining cohorts performed better than the reference group, and in one case was the difference statistically significant. Findings suggest that self-explaining activities support students' conceptual understanding at least as much as instruction. This study contributes evidence for the self-explaining effect and the ICAP hypothesis in a discipline where no evidence is available. Furthermore, it adds to the relatively little work in self-explaining that has explored naturalistic learning environments. This work supports the incorporation of self-explaining activities in the repertoire of instructional practices for General Chemistry. All Rights Reserved © 2015 Universidad Nacional Autónoma de México, Facultad de Química. This is an open access item distributed under the Creative Commons CC License BY-NC-ND 4.0.

* Corresponding author.

E-mail address: santiago.s.urena@gmail.com (S. Sandi-Urena).

¹ Current address: Department of Chemistry and Biochemistry, California State University-Bakersfield, 9001 Stockdale Highway, Bakersfield, CA 93311, USA.

² Current address: Escuela de Química, Universidad de Costa Rica, San Pedro de Montes de Oca, 2060 San José, Costa Rica.

Peer Review under the responsibility of Universidad Nacional Autónoma de México.

PALABRAS CLAVE

Auto-explicación;
Estudio en salón de
clase;
Educación química;
Explicaciones
científicas;
Química General

Evaluación del efecto de la auto-explicación en un curso de Química General de matrícula masiva

Resumen La auto-explicación se refiere a la generación de inferencias para el consumo propio sobre conexiones causales entre objetos y eventos. La auto-explicación es una de las prácticas de la ciencia que se consideran esenciales para la competencia científica; por tanto, en sí misma es un valioso producto de aprendizaje. Sin embargo, la generación de explicaciones auténticas en la instrucción universitaria de ciencias se promueve muy poco. Este trabajo examinó el efecto que generar auto-explicaciones tiene sobre la comprensión conceptual de la química. El aprendizaje y la ejecución de las tareas fueron completados individualmente en la ecología natural de un salón de clases en un curso masivo de Química General en los EEUU. El estudio se prolongó por cinco semestres incluyendo las pruebas piloto y las réplicas. La intervención de auto-explicación siguió un diseño comparativo de condiciones múltiples que usó un post-test para evaluar el aprendizaje. Los estudiantes fueron asignados aleatoriamente a las siguientes condiciones: revisión de una explicación correcta, explicación de respuestas correctas o incorrectas, explicación de la opinión sobre preguntas producidas por otros, explicación de las respuestas propias. Se usó de referencia un cohorte que recibió instrucción estándar sin intervención y que se había preparado para la examinación formal. Los cohortes de auto-explicación ejecutaron el post-test mejor que el grupo de referencia; en un caso la diferencia fue estadísticamente significativa. Los resultados sugieren que las actividades de auto-explicación apoyan la comprensión conceptual al menos tanto como la instrucción dirigida. Este estudio aporta evidencia para el efecto de auto-explicación y la hipótesis ICAP en una disciplina en que tal evidencia no estaba disponible. Más aún, suma al relativamente poco trabajo en auto-explicación que ha explorado ambientes de aprendizaje en la ecología natural del salón de clases. Este trabajo apoya la incorporación de auto-explicaciones en el repertorio de prácticas de aprendizaje en la enseñanza de Química General.

Derechos Reservados © 2015 Universidad Nacional Autónoma de México, Facultad de Química. Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons CC BY-NC-ND 4.0.

Introduction

As the volume of material assigned in entry-level college science courses continues to expand and class enrolment increases, the expectation that substantial learning will occur in the confines of the classroom becomes less tenable. Many first-year college students struggle with the expectation that most learning should occur away from instructor supervision (Conley, 2007). Additionally, evidence suggests learning outside the classroom tends to occur when the learners are unaccompanied (Villalta-Cerdas & Sandi-Urena, 2013). Therefore, it is reasonable to promote classroom-learning strategies transferable to *where and how* unsupervised learning occurs. In light of this, we investigate how to foster self-explaining—an individual learning strategy—in the chemistry classroom context (Villalta-Cerdas & Sandi-Urena, 2014). Furthermore, the US National Research Council has identified self-explaining as one of eight practices deemed essential and desirable for scientific competence (National Research Council, 2013). Besides being a valued learning outcome in itself, the ability to generate one's own explanations supports conceptual learning. Generating scientific explanations engages individuals in analysis and reflection of current models and theories, thereby developing their familiarity and proficiency, and influencing conceptual understanding (Taber & Watts, 2000). Science-learning strategies nurture conceptual understanding so learners can gain the knowledge necessary to solve problems effectively and efficiently. Ford

and Wargo (2012) postulate understanding a concept in science is both conceptual and epistemic in nature, and this understanding becomes evident in one's ability to use that concept in explanation and argumentation. Thus, it is through the generation of well-grounded explanations that meaningful understanding may be assessed (Talanquer, 2009). Not surprisingly, support for the beneficial role of constructing scientific explanations abounds, especially in the inquiry literature (Ryoo & Linn, 2014). Research has shown that environments that engage students in scientific explanations can enhance their knowledge, epistemic practices, and literacy skills (Ryoo & Linn, 2014). With few exceptions (Obaya Valdivia, 2004), chemistry education research on explanations has centered majorly on the nature of explanations and their descriptions (Stefani & Tsapartlis, 2009; Taber & Watts, 2000; Talanquer, 2009, 2013), often from the perspective of instruction (Talanquer, 2007; Treagust, Chittleborough, & Mamiala, 2003). In this study we explore student engagement in the process of generating authentic explanations, by and for themselves, through a General Chemistry in-class activity, and its impact on conceptual understanding.

Self-explaining

Self-explaining refers to individuals' generation of their own explanations for their own consumption. It is an individual, non-interactive process different from explaining for others,

co-explaining, and similar mechanisms effective to enhance understanding and task performance but that, unlike self-explaining, require some form of dialog (Hausman, Chi, & Roy, 2004; Sandi-Urena, Cooper, & Stevens, 2012). During self-explaining, individuals create inferences about causal connections between objects and events (Siegler & Lin, 2009), thus, “self-explaining is a knowledge-building activity that is generated by and directed to oneself” (Chi, 2000, p. 165). In science, we see self-explaining as making sense of how and why actual or hypothetical phenomena occur. Evidence from two decades of research consistently supports students learn better when they self-explain the materials they are studying (Chi & Wylie, 2014). This positive impact on learning—the self-explaining effect—has been widely replicated and shown to be domain-independent (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Hausmann & VanLehn, 2010; Wylie & Chi, 2014). The outputs produced through engagement in self-explaining are referred to as self-explanations. When the process of self-explaining is successful, the self-explanation is content-relevant and contains a new piece of knowledge generated through the process.

The Interactive-Constructive-Active-Passive, ICAP, theoretical framework (Chi, 2009) to which we ascribe in our work (Villalta-Cerdas & Sandi-Urena, 2014) advances the theoretical underpinnings that support the self-explaining effect (Chi & Wylie, 2014). ICAP classifies learning activities based on learner’s overt, externally observable behaviors (Chi, 2009). This taxonomy identifies four modes of student engagement in a learning event: interactive, constructive, active, and passive. In an *interactive* activity learners engage in exchanges with others (e.g. argumentation) where all participating individuals contribute substantively and collaboratively. The hallmark of *constructive* activities is the generation of an output (e.g. information, ideas) that goes beyond what was originally explicit in the task itself (e.g. an inference). *Active* refers to the physical engagement in the activity without a novel output or interaction, for instance, underlining text. Responses to a learning task not falling in one of these foregoing classifications are *passive*. Two fundamental pillars of ICAP are: (a) the taxonomy uses learners’ observable behaviors and not the activity’s intended purpose or the instructor’s involvement, and (b) higher modes of engagement subsume lower modes in the sequence Interactive-Constructive-Active-Passive. According to the latter, for example, an interactive activity contains all the characteristics of a constructive activity and, in addition, there are interpersonal exchanges. As a corollary, activities classified as constructive are necessarily individual. Naturally, individuals may engage in constructive behavior with others, however, in alignment with these operational definitions, such engagement would be *interactive*. An implication of this hierarchy particularly relevant here is that learning activities produce greater learning outcomes when they are interactive compared to constructive. Likewise, constructive activities are more efficient than active and active more than passive (Chi, 2009). This differentiated effectiveness in producing learning is the ICAP hypothesis (Chi, 2009; Menekse, Stump, Krause, & Chi, 2013) whose postulates have been validated empirically through extensive analysis of published studies in both research laboratory and classroom settings (Chi, 2009; Chi & Wylie, 2014).

On the other hand, outcome differences amongst learning tasks within the same ICAP mode (e.g. constructive) have been found to be minimal (Fonseca & Chi, 2011). It is noteworthy to reiterate that self-explaining is a constructive activity in the ICAP framework and as such individual (Chi, 2009; Villalta-Cerdas & Sandi-Urena, 2014).

We identified a void in domain-specific self-explaining research pertaining to chemical education. Furthermore, even STEM-focused research only rarely draws samples from STEM majors and typically uses laboratory settings, and not natural learning environments (Villalta-Cerdas & Sandi-Urena, 2013). Cognizant of how these factors might limit the instructional use of self-explaining by chemistry educators (Villalta-Cerdas & Sandi-Urena, 2013) and convinced that generating explanations is essential in the practice of the sciences (National Research Council, 2013), we endeavored to develop research that would contribute to filling the void. In an initial stage, we gathered evidence in large enrolment General Chemistry settings that supports an association between the way tasks are framed and engagement in self-explaining behaviors of different sophistication levels (Villalta-Cerdas & Sandi-Urena, 2014): Appropriately designed instruction can effectively modify self-explaining practices in naturalistic environments. Moreover, sustained prompting of self-explaining may result in this behavior becoming habitual and transferable. Empirical evidence gathered in classroom environments may encourage instructors to overcome the practical challenges faced when implementing educational research. A natural complement to our previous work is the investigation of the effect that self-explaining may have on conceptual learning in chemistry. The present work tackles this objective.

Methodology

Goals and research questions

The current report focuses on the assessment of the self-explaining effect (Chi et al., 1989) in the context of large enrolment General Chemistry: How self-explaining influences learning of specific chemistry content. Two major characteristics of this study are: (a) participants were in their normal student function and in their natural learning environment and (b) we focused on conceptual understanding rather than learning declarative or procedural knowledge (e.g. using worked-out examples).

The following questions guided our investigation:

- 1) Does engagement in self-explaining activities influence students’ understanding of chemical concepts? To address this question we compared performance on a post-learning task for students who participated in a self-explaining activity with that of students who underwent instruction without the activity.
- 2) Do tasks with differential demand of self-explaining engagement elicit differences in students’ conceptual understanding? To address this question we compared performance on a post-learning task for students in several self-explaining conditions.

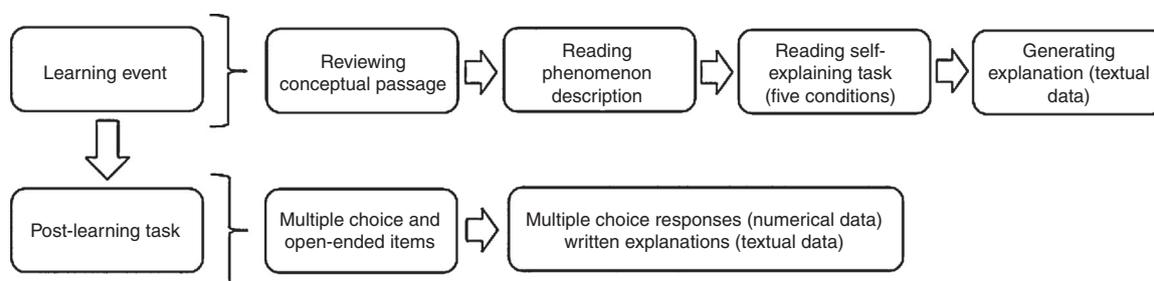


Figure 1 Administration of intervention materials.

Table 1 Intervention tasks (SE-Task).

Type of task	SE-Task	Description of prompt
No Self-explaining	NE	Reviewing correct explanation
Self-explaining	SEA	Explaining one's own answer
	EADA	Considering others' answers and explaining one's agreement/disagreement
	SEO	Explaining one's own answer for others to use while studying
	SEIA	Explaining others' incorrect answer and possible reasoning

Context and participants

This work used a naturalistic setting and data from General Chemistry 2 students at a large, urban, public, research university in the US. The institution served over 39,000 students (78% undergraduate); enrolment in General Chemistry sections was 200. Only sections taught by the corresponding author took part in this study. Participants were performing in their natural learning environment; therefore, we refer to them exclusively as *students*. Diverse ethnic minority students made up 39% of the undergraduate student body.

The tasks in this work were embedded in the course; therefore, from the students' perspectives they were simply part of ordinary assignments. The self-explaining activity, or intervention, was conducted before students were formally introduced to the concepts in the study. Grading guidelines were the same for similar assignments. Credit was received for satisfactory completion of the activity and not based on accuracy. We only used data from students who previously agreed to participate. Gender distribution of samples was representative of the university demographics.

Study design and materials

This project was conducted over five semesters to allow for pilot-tests and replications. The intervention followed a multi-condition comparison design that used post-learning knowledge performance to assess conceptual understanding. The domain included entropy and the Second Law of Thermodynamics, which we treated as individual knowledge components (Van Lehn, 2006). The intervention materials (available as Supplemental Materials³) were specifically

developed as research tools. A detailed description of their development and use as well as a thorough description of the larger research project are available elsewhere (Villalta-Cerdas, 2014).

Fig. 1 shows the administration design of the intervention. The learning event included a passage with a general description of the Second Law of Thermodynamics. Following, the learning event presented a familiar phenomenon common to all the conditions: Water freezes spontaneously below 0°C. Although familiar, this phenomenon challenges students' understanding of entropy and the Second Law of Thermodynamics. Next, each student had one of five different tasks, SE-Tasks (Table 1). One task, NE, instructed students to *review* a written explanation while the other four (SEA, EADA, SEO, and SEIA) required self-explaining. The latter four SE-Tasks prompted students to engage in explaining to reconcile the observation with the chemistry concepts. Each SE-Task defined a study condition (Table 1) and, by design, they were calibrated to promote different levels of self-explaining engagement. We based the calibration on literature reports (Fonseca & Chi, 2011), especially multi-condition comparison studies (Siegler & Lin, 2009). In the No self-explaining task (NE-Task) students reviewed an expert's explanation. Thus, the NE-Task did not require the construction of an externalized output and was—in principle—an ICAP passive activity. This condition functioned as a comparison against the other SE-Tasks—ICAP constructive activities—that prompted students to produce written explanations and, therefore, most likely to engage in self-explaining.

The learning event (~15 min) was immediately followed by the post-learning task (~10 min), a near transfer task in both content (what is transferred) and context (when and where transfer occurs), as opposed to a far transfer task where transfer occurs in a dissimilar context (Barnett & Ceci, 2002). Here students worked on five problems that prompted them to predict changes in entropy and to reason the spontaneity of processes. The post-learning task

³ Supplemental Materials are available directly from the Authors upon request (Research instruments, Assessment rubric for research materials, Validity and reliability evidence of post-learning task) and through *Educación Química*.

Table 2 Data collection by Study Phase.

Study phase		Data collection semester	Sample size (N)
Pilot Study	Pilot Study I	Fall 2011	130
	Pilot Study II	Spring 2012	443
Main Study		Fall 2012	174
Replication Study		Spring 2013	147
No Intervention		Fall 2013	142

is a criterion-referenced test, designed to determine students' application, analysis, and evaluation of the material introduced during the learning event. Of the five items, the first four were two-tiered and required the prediction and explanation of changes in entropy that involved the dissolution of an ionic compound in water, the addition of a gas to water, or folding of proteins. For example, in the scenario using proteins students first predicted the change in entropy of proteins folding into their native structures by selecting the correct statement in a multiple-choice question; then they explained their choice in writing. The written explanations (textual data) allowed insight into the students' self-explaining sophistication; their analysis is the matter of a separate report. The fifth item was a multiple-choice question to probe understanding of the Second Law of Thermodynamics and, unlike the other four, it was not two-tiered. The present study discusses the multiple-choice assessment (numerical data).

We divided the entire study into four phases, [Table 2](#): (1) Pilot Study (composed of two cohorts, Pilot Study I and Pilot Study II); (2) Main Study; (3) Replication Study; and (4) No Intervention. The intervention phases (Pilot, Main, and Replication Studies) followed the design described in [Fig. 1](#). Students in the No Intervention Phase completed the post-learning task only after ordinary course instruction and course assessment on entropy and the Second Law of Thermodynamics. Data gathering for the No Intervention phase occurred during the fifteenth week of the semester. In our view, these data are representative of what a cohort would normally learn in class with the same instructor and curriculum but excluding the intervention. Therefore, these data served as a baseline to assess conceptual understanding using the post-learning task, that is, the No Intervention group functioned as a reference.

The Pilot Study served to test logistics and gain insight about the efficacy of materials and procedures as well as data analysis ([van Teijlingen & Hundley, 2001](#)). Materials were reviewed and no major changes resulted from this process. We consulted and held meetings with three chemistry education doctoral candidates and two experienced chemistry instructors who offered general advice and completed assessment rubrics to evaluate content validity (Supplemental Materials). Finally, two external and experienced chemical education researchers, assessed the content and construct validity of the instruments independently. All reviewers agreed the instruments were adequate for the assessment of entropy and the Second Law of Thermodynamics ([Villalta-Cerdas, 2014](#)).

Validity evidence for the post-learning task is available in [Villalta-Cerdas, 2014](#), Supplemental Materials, and upon request from the Authors.

Data collection

Other than the self-explaining prompt in the learning event, the intervention materials were identical for all students. Materials were administered during class and without student interactions. Students worked on the learning event first, and only after they had turned it in did they have access to the post-learning task. Students started and turned in each task simultaneously. These procedures were consistent with class norms established previously. Multiple-choice responses for the post-learning task were collected using Optical Mark Recognition sheets (*scantrons*). Written explanations were collected and processed for further analysis ([Villalta-Cerdas & Sandi-Urena, 2014](#)). [Table 2](#) shows the sample sizes for data collection.

For the intervention phases—Pilot, Main, and Replication Studies—data gathering occurred the tenth week of the course and before entropy and Second Law of Thermodynamics were presented in class. For Pilot Study I, we distributed alternate forms of the five SE-Tasks ([Table 1](#)). In this pseudo-randomized procedure, the probability of assignment to a given SE-Task was not independent for each individual. In the case of Pilot Study II, instead of assigning conditions to individuals, we used General Chemistry Laboratory sections as unit of distribution so all students in a given section were in the same condition. Gathering data in the laboratory setting was a convenient way to increase the sample size for validity evidence purposes (Supplemental Materials). This was the sole purpose since our interest focused in the classroom environment.

To meet conditions for true randomization ([Cook & Payne, 2002](#); [U.S. Department of Education, 2003](#)) in further Study Phases, we assigned students to the SE-Tasks using random number generation ([Microsoft, 2010](#)). The Replication Study did not include the No Self-explaining task (NE-Task). In all phases of data collection we took measures to minimize potential threats to internal validity such as the Hawthorne Effect ([Franke & Kaul, 1978](#); [Jones, 1992](#)). This included following procedures (e.g. distribution of materials and delivery of instructions) that were not different from procedures used for in-class assignments. We assumed minimal researcher intrusion and student familiarity with these procedures prevented predisposition.

Data analysis

Post-learning task performance comparisons

We compared performance on the post-learning task by condition to investigate the self-explaining effect. When applicable, we used independent comparisons for each Study Phase, which allowed us to check reproducibility of

Table 3 Post-learning task performance comparison between No Intervention and self-explaining activities.

Study Phase	Comparison groups	<i>n</i>	<i>b</i> -Coefficient	<i>p</i> -Value	<i>R</i> ²
Pilot Study I	No Intervention	142	−.18	.05	.03
	Self-explaining activities	130			
Main Study	No Intervention	142	−0.20	.10	.04
	Self-explaining activities	168 ^a			
Replication Study	No Intervention	142	−0.12	.17	.02
	Self-explaining activities	145 ^b			

^a Six participants excluded due to unintelligible written responses during the learning event.

^b Two participants excluded due to missing data in all items of post-learning task.

findings. We carried out two sets of comparisons, each one to address one of the guiding research questions:

- 1) Comparisons between students in the self-explaining conditions and students who completed instruction without intervention (No Intervention Phase) produced evidence regarding the self-explaining effect. The Pilot Study I, Main Study, and Replication Study (Table 2) each produced a set of self-explaining conditions that we compared independently with the No Intervention group.
- 2) Comparisons between students in the different self-explaining conditions, SE-Tasks (i.e., NE, SEA, EADA, SEO, and SEIA; Table 1), allowed us to probe the effect that the prompt modifications in the different self-explaining engagement activities had on conceptual understanding.

We used structural equation modeling (SEM) to perform the comparison studies where the condition membership of students (SE-Task) served as independent categorical variable acting upon the latent factor of the post-learning task (Bagozzi & Yi, 1988; MacCallum & Austin, 2000; Muthén, 1984). The input variables were the multiple-choice items on the post-learning task. For condition membership we created a binary variable entering 1 for students who were members of the interest condition and 0 for those who were non-members. The result of the SEM analysis provided a *b*-coefficient (β) value for the regression of the independent categorical variable (condition membership) on the latent factor, the significance value (*p*-value) for the *b*-coefficient, and the *R*² of the regression equation (Wang & Wang, 2012). The *b*-coefficient (−1.0 to 1.0) can be interpreted as the performance advantage that members of the interest condition have over non-members of the condition. The *R*² is a measure of the total amount of variance explained by the regression equation, and it can be interpreted as how much of the variation around the mean is associated with condition membership (SE-Task). For the SEM analyses we used MPlus (Version 6 from Muthén & Muthén Copyright © 1998–2010 Muthén & Muthén).

Results and discussion

Performance comparisons between self-explaining activities and instruction without self-explaining

This section addresses our first guiding question: Does engagement in self-explaining activities support

students' understanding of chemical concepts? We performed separate analyses to compare the post-learning task performance of students who did not participate in the self-explaining activities, the No Intervention group, with the three independent groups of students in the self-explaining activities: Pilot Study I, Main Study, and Replication Study (Table 3). These results showed that the No Intervention group performed consistently lower than students who self-explained in the intervention phases. It stands out that the *b*-coefficient was negative for all comparisons indicating lower performance for the No Intervention group. The difference in performance between the No Intervention group and the Pilot Study I was statistically significant at 95% confidence level. Comparison with the Main Study and the Replication Study produced $p = .10$ and $p = .17$, respectively. In previous work (Sandi-Urena et al., 2012) we have underscored the pertinence of carefully planned replication in field studies to enhance internal validity of applied research (Schafer, 1999). We echo Bauernfeind's stance (1968) that when a variety of replication studies lead to similar results and consistent conclusions, the *burden of proof* shifts toward those who would dispute such conclusions. The three-stage design—Pilot, Main, and Replication Studies—strengthened the interpretation of the statistical results. Furthermore, “results that have been replicated are considered more likely to generalize (continue to be observed) to further replications” (Schafer, 1999, p. 148). Findings drawn from these comparisons are not only harmonious across the phases of this particular study, but they are also consonant with preeminent reports in the field (Chi, Leeuw, Chiu, & LaVancher, 1994; Chi, 2009).

Students who completed the self-explaining activities—those in the Pilot, Main, and Replication Studies—were not formally introduced to the concepts of entropy and the Second Law of Thermodynamics prior to participating in the activities. Therefore, they learnt new material on their own that we assessed shortly thereafter. It is paramount to point out again that the No Intervention group was assessed using the same materials but after instruction, and after the students had prepared for the exam including the same topics. We propose that by this time they had learnt as much as they would in this class. This maximized performance became the bar by which we assessed conceptual understanding. The repeated measures suggest there was no difference between the students who were induced to engage in self-explaining as a learning strategy and those who instead underwent instruction and prepared for formal assessment.

Table 4 Post-learning task performance comparison between No Self-explaining and Self-explaining Tasks.

Study Phase	Comparison groups	<i>n</i>	<i>b</i> -Coefficient	<i>p</i> -Value	<i>R</i> ²
Pilot Study I	No Self-explaining	27	0.13	.31	.02
	Self-explaining Tasks	103			
Main Study	No Self-explaining	40	−0.01	.90	.00
	Self-explaining Tasks	128 ^a			

^a Six participants excluded due to unintelligible written responses during the learning event.

These results (Table 3) support our hypothesis that eliciting self-explaining through classroom activities effectively supports chemistry conceptual understanding. This finding is in alignment with research that has shown the beneficial effect of engaging in self-explaining is applicable to non-procedural domains and achievable through prompting to self-explaining (Chi & Wylie, 2014; Chi et al., 1994). Moreover, this work puts forth unique evidence in naturalistic environments and in a discipline where there is an identified research gap (Villalta-Cerdas & Sandi-Urena, 2013). That students learn when they become “cognitively engaged with the learning materials, even with no expert present to teach, correct errors, or explain misconceptions” may seem rather intuitive (Wylie & Chi, 2014, p. 415). Nonetheless, we argue the opposite notion—that learning requires tightly controlled mediation by an instructor—still prevails amongst college science educators. Evidence generated in classroom environments contributes to dispel this persistent assumption and to modify concomitant instructional practices. Research has established that self-explaining can be successfully prompted and students trained instead of relying solely on spontaneous generation of self-explaining (Roy & Chi, 2005). Classroom training in effective individual learning strategies that are transferable to *where* and *when* students do their unsupervised learning will assist them in realizing their learning potential (Villalta-Cerdas & Sandi-Urena, 2013, 2014).

Performance comparisons between self-explaining tasks

In this section we focus on the second guiding question: Do tasks that demand different self-explaining engagement elicit differences in students’ conceptual understanding? Initially we hypothesized that tasks with different demand of self-explaining engagement could lead to differences in students’ conceptual understanding. We compared post-learning task performance for students in the SE-Tasks: NE, SEA, EADA, SEO, and SEIA. The SE-Task membership functioned as an independent, categorical variable acting upon the latent factor in the structural equation model for the post-learning task.

First, we investigated the differences between students in the No Self-explaining task and the students in all of the other self-explaining conditions combined. Unlike the No Intervention group, the No Self-explaining task condition participated in the intervention and completed the post-learning task without any instruction or additional preparation time. The No Self-explaining task directed students to *review* an explanation for the phenomenon in the

learning event instead of *generating* one; hence, learning in this condition relied—in principle—solely on expository text. On the contrary, the self-explaining conditions (i.e. SEA, EADA, SEO, and SEIA) required an externalized output (written text). Therefore, in our activity design we initially proposed the No Self-explaining task as a *passive* learning activity, whereas the others were *constructive*. This supported the rationale that the comparison was between the passive and constructive ICAP modes of engagement (Chi, 2009). Table 4 shows the results for this set of comparisons. There was no significant difference in performance between the No Self-explaining and self-explaining conditions for the Pilot Study I and the Main Study. In both cases, the *b*-coefficients in the SEM analyses were small and non-significant (Table 4). In light of these results, membership in either the No Self-explaining task or in a self-explaining condition (i.e. SEA, EADA, SEO, and SEIA) carried no significant additional benefit in performance.

In principle, this unexpected finding seemingly contradicted the ICAP hypothesis that constructive activities are more effective in producing learning than passive activities (Fonseca & Chi, 2011; Menekse et al., 2013). However, recent work (Chi & Wylie, 2014; Villalta-Cerdas & Sandi-Urena, 2014) has drawn attention to the difference between the intended overt behavior and that enacted by the students. The No Self-explaining task was not intended to induce any engagement other than passively reading the material. Nevertheless, further scrutiny of the intervention environment and students’ responses to this environment led us to the re-interpretation of the results. We came to the realization that inducing a purely passive behavior in this setting was, paradoxically, a challenge. All students participated in the activity simultaneously and in the same room. Students in the No Self-explaining condition could observe peers in other conditions overtly engaged in processing information and in writing responses. Proof to the effect of the surroundings is that despite the fact that the instructions did not require them to do so, a considerable number of students in the No Self-explaining condition opted to jot down their ideas, interpretations, or processing of the explanation presented to them. Furthermore, even when students in this condition restrained from creating the externalized output, this environmental effect was likely to stimulate covert engagement in cognitive processes associated with the overt behaviors of the self-explaining conditions. In other words, a misalignment between overt behavior and covert processing might have occurred (Chi & Wylie, 2014). That the outputs of the cognitive processes do not necessarily have to be externalized for self-explaining to occur is clear (Chi, 2009). However, from a researchers’ perspective the assessment of internal outputs is not practical, if

Table 5 Post-learning task performance comparisons across SE-Tasks by Study Phase.

SE-Task/Non SE-Tasks comparison	Study phase (<i>n</i> SE-Task/ <i>n</i> Non SE-Tasks)	<i>b</i> -Coefficient	<i>p</i> -Value	<i>R</i> ²
SEA-Task/Non SEA-Tasks	Pilot Study I (25/78)	-0.06	.64	.00
	Main Study ^a (29/99)	0.14	.22	.02
	Replication Study ^b (35/110)	-0.14	.21	.02
EADA-Task/Non EADA-Tasks	Pilot Study I (28/75)	-0.18	.27	.03
	Main Study ^a (31/97)	-0.02	.88	.01
	Replication Study ^b (37/108)	-0.10	.45	.00
SEO-Task/Non SEO-Tasks	Pilot Study I (23/80)	0.19	.33	.04
	Main Study ^a (35/93)	0.02	.89	.00
	Replication Study ^b (38/107)	-0.06	.69	.00
SEIA-Task/Non SEIA-Tasks	Pilot Study I (27/76)	0.05	.73	.00
	Main Study ^a (33/95)	-0.13	.30	.02
	Replication Study ^b (35/110)	0.15	.30	.02

^a Six participants excluded due to unintelligible written responses in learning activity.

^b Two participants excluded due to missing data in all items of post-learning task.

feasible at all. In support of this re-interpretation and in addition to the environmental influences, we argue that the instruction *to review* used in the No Self-explaining condition might have not constrained students to exclusively behave passively. *To review* may be interpreted as critically examining, reflecting, or evaluating. Chi (2009) described a similar occurrence in which, upon closer inspection, a study condition that was presumably *active only* turned out to be constructive. Somewhat analogous to our case, nuances in the instructions for that condition, although not requiring an observable output, unintentionally led participants to engage in deep reflection of a phenomenon. Consequently, the output of self-explaining remained non-observable. We concluded that our intention to implement a passive condition was a failed attempt: Environmental cues in conjunction with the apparent openness of the language in the instructions prompted students in the No Self-explaining condition to engage in cognitive processes consistent with self-explaining. In light of this re-interpretation, results in Table 4 fit with the ICAP prediction that learning activities of the same engagement mode produce non-significant differences in learning outcomes. Note again that the ICAP framework makes claims based only on overt engagement and cognitive processes that an observer can relate to students' behaviors (Menekse et al., 2013). That is, the ICAP taxonomy is based on what the learners do with the activity not what the instructor does or intends to happen (Chi & Wylie, 2014). Although students can covertly engage in all kinds of sophisticated cognitive processes, the ICAP premise is that certain processes are more likely to accompany certain overt behaviors (Fonseca & Chi, 2011; Menekse et al., 2013). Once we realized the No Self-explaining condition was not fulfilling the design purpose, we decided not to use it in the Replicate Study since it was not possible to isolate participants from the environmental influences without removing them from the natural environment.

Next, we compared post-learning task performance across the constructive SE-Tasks: SEA, EADA, SEO and SEIA. This was a comparison within a single ICAP engagement mode (constructive) with variations in the learning activity (self-explaining). For this purpose, we compared membership in a given condition (for instance, SEA-Task) with no

membership in that condition (Non SEA-Task), that is, the difference between participating in one condition (SEA-Task) versus participating in any of the other three (EADA, SEO, and SEIA). To assess reproducibility, we performed these comparisons separately for each Study Phase. Table 5 shows the four possible comparisons of each SE-Task with the Non SE-Tasks group (first column) and the number of students in each of these groups for each Study Phase (second column). There were no statistically significant differences in performance as a function of SE-Task membership. For all study phases, we observed non-significant, low *b*-coefficient values (range: -0.2 to 0.2) with no discernable trend across the phases.

Our initial hypothesis adhered to the stance that amount of self-explaining and amount of learning are correlated (Siegler, 2002). It was also based on the assumption that instructors can "design tasks that elicit more or less engagement from students" (Chi & Wylie, 2014, p. 219); therefore, differences in prompts could translate into different amounts of self-explaining and thereby produce different learning performances. Evidence has shown, for instance, that middle school students that generated a large number of self-explanations developed a better understanding of the circulatory system than those who generated fewer self-explanations (Chi et al., 1994). However, results from Table 5 suggest that the self-explaining effect on our students' conceptual learning was, for all purposes, independent of the prompt used in the individual conditions. By all appearances, students in these conditions engaged to a similar extent in the underlying cognitive processes that mediate learning and are associated with constructive activities in general and self-explaining in specific (Chi, 2009; Chi & Wylie, 2014). With the benefit of hindsight, we argue that the self-explaining tasks were narrowly delimited for the students and did not lend themselves to engagement in considerably different amounts of self-explaining. On the other hand, results in Table 5 are in alignment with predictions of learning outcomes based on the ICAP framework. Chi (2009) and collaborators (Fonseca & Chi, 2011) have applied the ICAP model to classify studies in the literature and have found that, under similar conditions, learning activities of the same ICAP engagement mode achieved

similar learning outcomes. Similarly, Nokes-Malach, VanLehn, Belenky, Lichtenstein, and Cox (2013) found no significant differences in far transfer test performance for students who learned physics from worked-out problems in one of two conditions using each a constructive activity: analogical comparisons or self-explaining. Nonetheless, in cases where the complexity of tasks allows learning activities within a mode to differ substantially they may lead to different learning outcomes (Chi & Wylie, 2014, p. 236). In the present study, all the SE-Tasks were constructive and, seemingly, differences in their framing did not reach the threshold to produce observable differences in learning outcomes. Our findings further support the ICAP framework hypothesis of expected learning outcomes based on students' overt behaviors when engaged in in-class activities.

Limitations

Albeit randomized, conditions originated from convenience samples: Students were all enrolled in the corresponding author's course. There is no practical way to determine whether particular factors influenced students' choice of section. If that were the case, samples would be a sub-set of the population. Although the instructor, instruction, and curriculum were not altered, data for the No Intervention comparison were collected in a semester different from the other phases. This limitation derives from practical reasons and may not be feasible to modify.

Finally, we believe that positive far transfer plays a major role in establishing the effectiveness of instructional interventions and the assessment of research (Bransford & Schwartz, 1999). Nonetheless, in the present study we are concerned with students' responses to an immediate environmental demand. Carrying out further observations at a later time during the semester may not produce relevant information. Even after a short time, confounding factors might obscure the effects elicited by condition membership. Students would have participated in other learning activities over which the research team exerts no control (e.g. lecture, study, and exam preparation). Near transfer fits our purpose to demonstrate that what is learnt through self-explaining may be used in the context of the classroom and course.

Conclusions and implications

Our previous work provided evidence that suggests it is possible to effectively promote engagement in self-explaining in the naturalistic environment of large enrolment General Chemistry (Villalta-Cerdas & Sandi-Urena, 2014). The significance of this finding is underscored by the observation that when left to their own devices students do not tend to spontaneously generate self-explanations (Woloshyn & Gallagher, 2009). Inclination toward paraphrasing may be the natural response to exposure to rhetoric of conclusions as instructional strategy (Schwab & Brandwein, 1962). To extend that work, we investigated the self-explaining effect in the same learning ecology. The present report supplements our previous findings by contributing evidence that links self-explaining with effective learning. Upon participation in a single self-explaining instantiation, students demonstrated

having learnt on their own by performing as well as peers who received instruction and had prepared for an exam. These results were consistent for self-explaining students in three independent cohorts in different semesters. This work supports that college students can self-explain chemical phenomena and develop conceptual understanding without direct instruction, and then transfer the underlying chemistry concepts to novel contexts. Moreover, there is an added value in the consistent use of self-explaining as a classroom instructional strategy: it fosters its adoption as a habitual learning strategy and science practice (Villalta-Cerdas & Sandi-Urena, 2014) as it improves over time (Wylie & Chi, 2014). While self-explaining is a cognitively demanding activity—which may impede spontaneous engagement (Woloshyn & Gallagher, 2009)—research evidence supports self-explaining is teachable and learners can be trained effectively to self-explain (Roy & Chi, 2005). Developing this learning strategy in the classroom will facilitate students' learning *when* and *where* they are not under direct supervision of an instructor and in the absence of peers (Villalta-Cerdas & Sandi-Urena, 2013).

As course contents and class enrolment increase, the expectation that substantial learning will happen in the classroom becomes more unrealistic. Instructors must contemplate supporting learning strategies consistent with students' needs. Approaches that require learning outside the classroom will not succeed if students are not equipped with the tools to advance autonomous learning. We believe that both less and more experienced instructors can effectively design and incorporate self-explaining activities in their repertoire of teaching practices (Villalta-Cerdas & Sandi-Urena, 2013). Entwistle and McCune (2004) argued 'the link between teaching methods and study strategies has been demonstrated, indicating the indirect influences that faculty members have on students' study behavior'. Thus by using classroom self-explaining activities, college instructors can address conceptual understanding while having an impact on students' later study habits. As a subject independent learning strategy, self-explaining has the potential to continue influencing learning beyond the boundaries of a single discipline and formal schooling.

In our work, we used tasks of the same ICAP mode—constructive—and the same learning strategy—self-explaining—in a design that utilized replication and random assignment with concurrent conditions. Consistent across three independent samples, performance in this study was not significantly different for students in the four self-explaining tasks. That is, all tasks seemed to positively impact conceptual understanding to a similar extent. Based on thorough re-analysis and interpretation of published results, Chi (2009) and Fonseca and Chi (2011) concluded that interventions of the same ICAP mode would produce no significant differences in learning outcomes.

Research suggests that one difference between better and worse learners is the extent to which they engage in attempts to explain what they are learning (Siegler, 2002). It also suggests that frequency of explaining is related positively to learning for both high ability and low ability learners (Chi et al., 1989). It follows then that all learners have the potential of becoming *better* learners by developing the habit to explain while learning and, given that most learning happens individually (Villalta-Cerdas &

Sandi-Urena, 2014), developing the habit of self-explaining. Like that of others' (Atkinson, Renkl, & Merrill, 2003; Chi, 2000; Songer & Gotwals, 2012), our work supports the claim that self-explaining activities facilitate learning in the sciences. However, in this research project we have explicitly addressed a gap in research, namely, the study of the self-explaining effect in college sciences in general, and in college chemistry in specific, and the need for evidence gathered in naturalistic learning environments. We trust college science educators will consider this supporting evidence when deciding whether to integrate self-explaining activities to the repertoire of their instructional design.

Conflict of interests

The authors declare no conflict of interest.

References

- Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*(4), 774–783.
- Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science, 16*(1), 74–94.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612.
- Bauernfeind, R. H. (1968). The need for replication in educational research. *Phi Delta Kappan, 126*–128.
- Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 61*–100.
- Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. *Advances in Instructional Psychology, 5*, 161–238.
- Chi, M. T. H. (2009). Active–constructive–interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73–105.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*(2), 145–182.
- Chi, M. T., Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439–477.
- Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243.
- Conley, D. T. (2007). *Redefining college readiness* (Vol. 3) Eugene, USA: Educational Policy Improvement Center.
- Cook, T. D., & Payne, M. R. (2002). Objecting to the objections to using random assignment in educational research. *Evidence Matters: Randomized Trials in Education Research, 150*–178.
- Entwistle, N., & McCune, V. (2004). The conceptual bases of study strategy inventories. *Educational Psychology Review, 16*(4), 325–345.
- Fonseca, B. A., & Chi, M. T. H. (2011). Instruction based on self-explanation. In R. E. Mayer, & P. A. Alexander (Eds.), *The handbook of research on learning and instruction* (pp. 296–321). New York, USA: Routledge.
- Ford, M. J., & Wargo, B. M. (2012). Dialogic framing of scientific content for conceptual and epistemic understanding. *Science Education, 96*(3), 369–391.
- Franke, R. H., & Kaul, J. D. (1978). The Hawthorne experiments: First statistical interpretation. *American Sociological Review, 623*–643.
- Hausmann, R. G. M., Chi, M. T. H., & Roy, M. (2004, August). *Learning from collaborative problem solving: An analysis of three hypothesized mechanisms*. pp. 547–552. Chicago, IL: Paper presented at the proceedings of the 26th annual conference of the Cognitive Science Society.
- Hausmann, R. G., & VanLehn, K. (2010). The effect of self-explaining on robust learning. *International Journal of Artificial Intelligence in Education, 20*(4), 303–332.
- Jones, S. R. (1992). Was there a Hawthorne effect? *American Journal of Sociology, 451*–468.
- MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51*(1), 201–226.
- Menekse, M., Stump, G. S., Krause, S., & Chi, M. T. (2013). Differentiated overt learning activities for effective instruction in engineering classrooms. *Journal of Engineering Education, 102*(3), 346–374.
- Microsoft. (2010). *Microsoft Excel [computer software]*. Redmond, Washington: Microsoft.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*(1), 115–132.
- Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.
- National Research Council. (2013). *Next generation science standards: For states, by states*. Washington, DC, USA: The National Academies Press.
- Nokes-Malach, T. J., VanLehn, K., Belenky, D. M., Lichtenstein, M., & Cox, G. (2013). Coordinating principles and examples through analogy and self-explanation. *European Journal of Psychology of Education, 28*(4), 1237–1263.
- Obaya Valdivia, A. (2004). Free energy and spontaneity through students' explanations and integration of their ideas. *Educación Química, 15*(4), 436–440.
- Roy, M., & Chi, M. T. (2005). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 271–286). New York, USA: Cambridge University Press.
- Ryoo, K., & Linn, M. C. (2014). Designing guidance for interpreting dynamic visualizations: Generating versus reading explanations. *Journal of Research in Science Teaching, 51*(2), 147–174.
- Sandi-Urena, S., Cooper, M., & Stevens, R. (2012). Effect of cooperative problem-based lab instruction on metacognition and problem-solving skills. *Journal of Chemical Education, 89*(6), 700–706.
- Schafer, W. D. (1999). An overview of meta-analysis. *Measurement and Evaluation in Counseling and Development, 32*(1), 43–61.
- Schwab, J. J., & Brandwein, P. F. (1962). The teaching of science as enquiry. In J. J. Schwab, & P. F. Brandwein (Eds.), *The teaching of science* (pp. 3–103). Cambridge: Harvard University Press.
- Siegler, S. (2002). Microgenetic studies of self-explanation. In N. Granott, & J. Parziale (Eds.), *Microdevelopment transition processes in development and learning* (pp. 31–58). New York, USA: Cambridge University Press.
- Siegler, R., & Lin, X. (2009). Self-explanations promote children's learning. In H. S. Waters, & W. Schneider (Eds.), *Metacognition, strategy use, and instruction* (pp. 85–112). New York, USA: Guilford Press.
- Stefani, C., & Tsaparlis, G. (2009). Students' levels of explanations, models, and misconceptions in basic quantum chemistry: A phenomenographic study. *Journal of Research in Science Teaching, 46*(5), 520–536.
- Songer, N. B., & Gotwals, A. W. (2012). Guiding explanation construction by children at the entry points of learning progressions. *Journal of Research in Science Teaching, 49*(2), 141–165.

- Taber, K. S., & Watts, M. (2000). Learners' explanations for chemical phenomena. *Chemistry Education Research and Practice*, 1(3), 329–353.
- Talanquer, V. (2007). Explanations and teleology in chemistry education. *International Journal of Science Education*, 29(7), 853–870.
- Talanquer, V. (2009). On cognitive constraints and learning progressions: The case of "structure of matter". *International Journal of Science Education*, 31(15), 2123–2136.
- Talanquer, V. (2013). When atoms want. *Journal of Chemical Education*, 90(11), 1419–1424.
- Treagust, D., Chittleborough, G., & Mamiala, T. (2003). The role of submicroscopic and symbolic representations in chemical explanations. *International Journal of Science Education*, 25(11), 1353–1368.
- U.S. Department of Education. (2003). *Random assignment in program evaluation and intervention research: Questions and answers*. Retrieved from: <https://www2.ed.gov/rschstat/eval/resources/randomqa.html>
- van Teijlingen, E., & Hundley, V. (2001). The importance of pilot studies. *Social Research Update*, 35, 1–4.
- Van Lehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, 16(3), 227–265.
- Villalta-Cerdas, A. (2014). *Development and assessment of self-explaining skills in college chemistry instruction* (Doctoral dissertation). Tampa: University of South Florida.
- Villalta-Cerdas, A., & Sandi-Urena, S. (2013). Self-explaining and its use in college chemistry instruction. *Educación Química*, 24(4), 431–438.
- Villalta-Cerdas, A., & Sandi-Urena, S. (2014). Self-explaining effect in general chemistry instruction: Eliciting overt categorical behaviours by design. *Chemistry Education Research and Practice*, 15, 330–340.
- Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus*. Hoboken, USA: John Wiley & Sons.
- Woloshyn, V., & Gallagher, T. (2009). *Self-explanation*. Retrieved from: <https://www.education.com/reference/article/self-explanation>
- Wylie, R., & Chi, M. T. (2014). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 413–443).