Learning by Arguing

Kalypso Iordanou[a], Deanna Kuhn[b], Flora Matos[b], Yuchen Shi[b], and Laura Hemberger[b]

[a]School of Sciences, UCLan Cyprus
[b]Teachers College, Columbia University, USA

Acknowledgment: Part of this work was funded by a grant from the Republic of Cyprus through the Research Promotion Foundation (Project: ΚΟΥΛΤΟΥΡΑ/BP-NE/0415/13) to Kalypso Iordanou.

Corresponding author:
Kalypso Iordanou, Ph.D.
UCLan Cyprus
12-14 University Avenue, Pyla, 7080 Larnaka, Cyprus
Tel: +357 24 81 21 21
Fax: +357 24 81 21 20
Email: KIordanou@uclan.ac.uk

Learning by Arguing

Abstract

Can argumentation practice simultaneously promote knowledge acquisition while advancing skill in the practice itself? We examine the effectiveness of a dialog-based argument curriculum in fostering middle-school students' knowledge acquisition as well as dialogic and written argumentation skill with respect to a content-rich, socially significant topic. Results of two studies, one involving a physical science topic and the other a social topic, showed a single intervention could meet both objectives. Study 1 following a previously used model of extended intervention (nine sessions over five weeks) and Study 2 experimenting with a dense model of six intervention hours completed over two days. Both were found effective. A novel question-and-answer method was found superior to a traditional one in promoting acquisition of factual knowledge sufficient to support argumentation. With respect to skill gains, a prompt to consider incongruent evidence showed the greatest effect in furthering mastery of a critical argument skill – to acknowledge and address, rather than ignore, evidence that counters one's favored position.

Key words: argumentation, dialog, discourse, writing, knowledge, learning

Learning by Arguing

Argumentation research has assumed a central place in recent years in both curriculum development and educational research. Research, however, has proceeded along two at most loosely connected fronts – arguing to learn and learning to argue. Those in the arguing to learn group are devoted to the idea that engaging in argumentation with peers is a promising path to students' knowledge acquisition (Andriessen & Baker, 2014; Asterhan & Schwarz, 2016; Weinberger & Fischer, 2006). Those in the learning to argue group focus more on argumentation as a valuable practice in its own right, with engagement and practice enhancing students' skills in both dialogic and individual written or verbal argument (Resnick, Asterhan, & Clarke, 2015; Kuhn et al., 2016a).

A major objective of the work presented here is to examine the extent to which both objectives can be achieved in the same set of activities. The question is an important one, more significant than simply that of whether efficiency can be achieved by incorporating two goals into one activity. In broad terms, it is a question that has been asked repeatedly in different ways going back as far as Dewey (1916) but one that continues to be heard today. How can one function as an autonomous thinker and learner while availing oneself of the abundant knowledge and expertise available from others? "The fundamental dilemma of education," in the words of Elmore (2018), " . . . is how to introduce learners to complex bodies of knowledge and expertise while at the same time placing them in the position of assuming responsibility and control over the process of their own learning" (p. 141) – the latter a condition essential to achieving skill in argumentive discourse. Barzilai and Chinn (2017) express the dual goal as one of "…fostering learners' capacity for autonomous

thinking and of promoting acquisition of knowledge that was created by communities of expert inquirers" (p. 20).

Evidence of gains in argumentation skill via engagement and practice are wide-ranging (see Resnick, Asterhan, & Clarke, 2015, for review). Does knowledge gain go hand in hand with argumentation? Can it even be clearly documented as an outcome of argumentation? A number of authors take a positive view (Bereiter & Scardamalia, 2018; Wegerif, 2012; Weinberger & Fischer, 2006). Yet reviews by Asterhan and Schwarz (2016) and by Van der Veen and Van Oers (2017) indicate that evidence of a causal link between argumentation and domain-specific knowledge gains is scant, a conclusion supported in a meta-analysis by Wecker and Fischer (2014). In their review, Asterhan and Schwarz focus on conceptual change as the desired outcome and see as central to success the eliciting of a variety of different opinions, solutions and viewpoints in an open discussion space and fueling dialog by highlighting these differences. As students seek to reconcile the differences in discussion, conceptual change stands to occur.

In many cases, however, students come to a topic without sufficient prior knowledge or ideas to make discussion productive. Many such topics are important ones that teachers would like their students to be informed about. We propose, then, that integrating skill and knowledge gains in a single intervention may be best approached as a three-part process. To make discourse on such topics productive or even feasible, we would claim, students must come into contact with new information regarding the topic in a way that makes it comprehensible to them (Murphy, Greene, Firetto et al., 2018). Additionally, they must integrate such information into an existing knowledge base if it is to be accessible and useful. Finally, they may then potentially make use of this new knowledge in an appropriate,

functional way in discourse or in individual argument. It can then be examined how such discourse further solidifies, organizes, or augments such knowledge, whether or not such changes are regarded as significant enough to qualify as conceptual change (von Aufschnaiter, Erduran, Osborne, & Simon, 2008; Zohar & Nemet, 2002).

It is the first component of this process that we focus on in Study 1. Can students learn effectively what they need to know about such topics in the context of a dialogic method that at the same time advances the goal of developing their skills of argumentive discourse and, ultimately, argumentive writing? In Study 2 we focus the majority of our attention on documenting students' acquisition of knowledge regarding the topic. To enhance generalizability, the topics students address encompass both physical and social science, a physical science topic in Study 1 (fossil fuel vs. solar energy) and a social topic in Study 2 (social security program vs. individual saving for retirement), both topics of contemporary significance and ones important for young people to engage with from the standpoints of both the factual knowledge and the social policy issues involved.

1. Study 1

How might students best acquire the knowledge they require in order to address unfamiliar, heavily knowledge-dependent topics? The most straightforward and a commonly used instructional method is by means of an introductory passage they are asked to read "as background information." A different method, which we have used in the previously cited studies as a means of enhancing the knowledge base students have available to bring to bear on their argumentation, consists of a set of short questions and answers (Q&A), ideally with at least some of the questions posed by the students themselves (Kuhn et al., 2016a). The superiority of this method over a more traditional one we have not previously put to rigorous

empirical test. Does information become of greater potential use if a question it answers has been posed? This is the research question we address in Study 1, as a foundation for the investigation in Study 2 of the knowledge acquisition that actually occurs during argumentation**.** The rationale underlying an affirmative answer to this question is that students will better appreciate and be able to make use of this information if they first recognize its potential argumentive purpose (in supporting or weakening a claim). In short, the knowledge becomes situated in a framework of the arguments it has the potential to serve.

1.1. Method

The Study 1 research design compares two conditions, identical except for the manner in which students acquire factual information regarding the topic. In the experimental (Q&A) condition, this occurred in the manner used in previous studies employing our method: It was presented in the form of a set of short questions and their accompanying answers, made available over time during the course of students' engagement with the topic, and offered with the suggestion that having an answer to this question might be useful in their argumentation. In a control (text) condition, the identical information was made available but as a traditional informational text rather than in Q&A format, presented at the beginning of the intervention as background information. The control condition, note, thus subtracts several features of the method of information acquisition that we hypothesize to be most effective. Rather than segmented, information is presented as a traditional single reading passage; it is not structured in the form of answers to questions students may have, and is presented initially all at once, rather than made available gradually over time. All of these are features we hypothesize will advantage potential use of this information on the part

of students in the Q&A condition, relative to the control condition. If our hypothesis is supported, further studies will be indicated to identify which of these individual features are critical to effectiveness.

1.1.1. Participants

Participants were 88 fifth and sixth graders (45 female) from a public non-selective elementary school in a suburban area of **Cyprus**. Students were mostly from middle-class families and within an average range of ability and academic achievement typical for this school. Half were 10-year-old fifth graders, and half were 11-12-year-old sixth graders. Students in both grades participated as part of their science classes, taught by the same teacher.

1.1.2. Procedure

The procedure followed that first reported on by Kuhn et al. (2008) and Kuhn and Crowell (2011), with some modest variations. Prior to the beginning of the intervention, all students were given an in-class assignment to write an essay on the topic of alternative sources for generating electricity in their country – natural gas vs. solar energy. The instruction was as follows: "Write the argument you would make to convince someone who didn't agree that your position is the better one. Make the best argument you can." Students were given as much time as they needed to complete their essays. The same essay assignment was repeated shortly after the intervention was completed.

The intervention procedure followed that implemented in previous work (Kuhn et al., 2016a; Iordanou & Constantinou, 2015). It took place during nine ninety-minute sessions occurring twice per week over a period of five weeks. The intervention was identical in all respects but one. Half of the fifth graders and half of the sixth graders were chosen randomly

to be in the *Text* condition described above, in which the information made available to them with respect to the topic took the form of a text suitable for the reading level of students of this age. In the *Q&A* condition, the identical information was made available but in brief pieces in question-and-answer format and provided over time during the intervention. By the end of the intervention students in the Q&A condition had available the same amount of information as students in the Text condition. (The school's division of the fifth grade and sixth grade classes into two equivalent sections who met separately made this manipulation possible.) In both conditions, students were encouraged to ask additional questions of their own and aided in securing answers, and all information once made available remained so throughout engagement with the topic. The specific information made available to students appears in the online Appendix A in each of the two formats.

The intervention sessions took place as follows.

Sessions 1-2. After a brief introduction by the teacher, students chose the alternative they preferred (gas v. solar) and assembled into same-side teams of 5-6 students each. The assignment was to generate reasons supporting their team's side, record them on individual cards, and share them with one another. During Session 2, the same small groups reflected on their reasons, consolidating and eliminating duplicates, and then ranked the reasons cards with respect to their strength.

Sessions 3-8. Same-side pairs were formed who remained together throughout these sessions. At the beginning of Session 3, students in the Text condition were presented a text of 1423 words (see online Appendix A) and allowed time to read it. This text then remained available for them to consult throughout Sessions 3-8. Students in the Q&A condition were presented a series of cards that taken together contained the identical information as appeared in the full

text. Each card contained a question on one side and the answer to the question on the other side[1].The student pair was asked to together choose four of these questions that it would be helpful to learn the answers to and were allowed to turn over these four cards to see the answers. At the next session the pair chose a new set of four cards, while the cards that they had chosen in previous sessions remained available to them.

At each session, the pair engaged in an electronic dialog via an instant-messaging platform with an opposing-side pair. After completing the dialog, the pair reflected on an electronic transcript of their dialog, with the support of two electronic reflection sheets, the ‘‘Other Argument’’ and the ‘‘Own Argument’’ reflections. With the help of the ‘‘Other Argument” Reflection Sheet, the pair’s task was to analyze the opposing side’s argument and reflect on the effectiveness of the counterargument they made and consider possible improvements to this counterargument. With the help of the “Own Argument” Reflection Sheet, the pair’s task was to review and evaluate the counterarguments made by the opposing side to their own arguments and their rebuttals to these counterarguments, and consider possible improvements to their rebuttals. If pairs finished with time remaining, they exchanged transcripts and reflection sheets with other pairs to give and receive feedback. At the next session, they engaged with a different opposing-side pair.

Session 9. In this final session, same-side groups reassembled and worked together to prepare for an ensuing whole-class debate. The group reviewed and reflected on the reflection sheets they had completed in previous sessions to decide what arguments to use in the debate. An adult coach facilitated these discussions. Students were encouraged to use different colored cards to summarize arguments, counterarguments and rebuttals and to include evidence to support each.

One week after completion of the intervention, students were asked to write a final individual essay on the topic. Instructions were the same as for the initial essay.

The first author prepared detailed lesson plans provided to the teacher and had several meetings with the teacher before the enactment of the curriculum as well as before and after each session to ensure that the intervention would be administered as intended. Furthermore, the first author observed a third of the sessions conducted. In addition to the first author, a colleague not involved in the work observed a number of the lessons, which were also audio recorded to monitor treatment fidelity. The author and colleague coded for the presence or absence of each of the components described in the lesson plans, with almost perfect inter-rater reliability. There were only minor deviations from the lesson plan in time spent on certain activities. The categories that were coded to determine the degree of implementation fidelity were the following: generating reasons, receiving information, engaging in electronic dialogs, completing reflection sheets, and preparing for whole-class debate. Author's and colleague's observations and recordings confirmed that the teacher adhered to the protocol, including all the components that were described in the lesson plans.

1.2. Results

The main data analyses compare the essays students wrote before and after the intervention. Due to student absences, a few students failed to write an essay at either initial or final assessment. At initial assessment, two students from the Text condition and one student from the Q&A condition were absent. At the final assessment, four students from the Text condition and three from the Q&A condition were absent. Multiple imputation was used in the analysis to deal with the missing data (Enders, 2010).

Overall, students showed good ability to make use of the available evidence in their essays. Students in both conditions showed considerable variation in their selection of evidence to include in their essays. Of the 24 pieces of evidence, only two failed to be included in an essay by one or more students.

1.2.1. Coding of essays

It should be noted first that the initial and final essays were students' only individual work. All other activity was in one or another collaborative form that we regarded as the most effective and therefore essential form in which to conduct the intervention. A student's final essay thus can be seen as a representation, albeit possibly incomplete, of what he or she had taken away from this extended collaborative engagement with the topic. Within the collaborative intervention, the student likely brought ideas to the topic, acquired others through peer interaction, and co-constructed others, in some combination that is worthy of close analysis. Nonetheless, it is possible to regard the essay as a representation of this student's own understanding at this point in time – both in terms of the topic knowledge, which we go on to examine in Study 2, and in terms of this student's own judgment of what arguments warrant making with regard to the topic. The latter we see as particularly important. The essay may of course underestimate the full extent of the information a student has gleaned related to the topic; yet it reflects all the student sees as useful in the context of the task at hand of constructing an argument on the topic.

For analysis, each initial and final essay was divided into idea units. One of the authors and another coder, blind to condition and time, segmented into idea units the 166 initial and final essays students wrote. Inter-rater reliability on segmenting was achieved on a subset of 25% of units with 90% agreement, and the first author proceeded with segmenting

the remaining essays, again blind to condition and time. Only segments that included a claim and evidence (evidence-based units) served as the data base for further analyses.

Sometimes, a student merely inserted mention of a piece of evidence in the essay without connecting it to a specific claim or to any broader argument being made.

If the connection between the cited evidence and the claim was not made adequately enough to be identifiable, the unit was coded as non-functional. If the unit included a claim and a supporting (or weakening) piece of evidence connected to it, it was coded as a functional unit. Of the total number of units coded, 67.07% were classified as functional units – 68.23% in the Q&A group and 66.01% in the Text group.

Each functional unit was further categorized according to the type of function it served. The coding system used is the one reported by Kuhn et al., (2016) and subsequently by others, based on the type of function the unit serves. Four functions can be identified:

Support my own. A statement serving to support one's own position (M+)

Weaken other. A statement serving to critique and thereby weaken the opposing position (O–)

Support other. A statement serving to acknowledge strengths of the opposing position (O+)

Weaken my own. A statement serving to acknowledge weaknesses of one's own position (M–)

These types have been shown to form a sequence of emergence during the course of students' engagement in argumentive discourse and writing over an extended period of an academic year or more. Students' arguments are first concentrated on serving to support their own position. Attention to the opposing position appears later, and initially these arguments

serve only the function of seeking to weaken it. Last to appear are statements that acknowledge merits of the opposing side, and finally, ones that identify weaknesses of one's own side – a theoretically plausible sequence given that the latter two types are not congruent with the arguer's position and therefore need to somehow be reconciled with it – an accomplishment not initially met when these types first appear (Kuhn et al., 2016). We give these latter two types particular attention in Study 2.

Inter-rater reliability on identification and coding of functional units in the present data was achieved on a subset of 25% of units. Cohen's Kappa was .81, indicating good reliability; differences were resolved by discussion and one of the authors proceeded with segmenting the remaining essays, again blind to condition and time.

1.2.2. Essay performance at initial and final assessments

Data were examined for outliers and these were ruled out (see online Appendix C). We first compared the two groups' performance at initial assessment to ensure that students in the Text and Q&A conditions performed equivalently. Independent sample t-tests applied in multiple imputed datasets showed that there was no significant difference between the Text and Q&A groups in the number of coded units, $M = 1.92$ and $M = 1.70$, respectively, $p = .483$, in the number of evidence-based segments, $M = 1.35$ and $M = 1.13$, respectively, $p = .318$, or in the number of units coded as functional, $M = 1.28$ and .82, $p = .076$, confirming that the two groups did not differ significantly at the outset.

Did the essays of the two groups differ in length? To assess the effects of group and time on the mean number of units produced in students' essays, a generalized linear mixed model (GLMM) was used. A GLMM showed that the overall model was significant,

$F(3, 172) = 33.998$, $p < .001$. The interaction between group and time was not statistically significant, $F(1, 172) = .387$, $p = .534$. The fixed effect of group was not statistically significant, $F(1, 172) = 1.530$, $p = .218$. The fixed effect of time was statistically significant, $F(1, 172) = 100.443$, $p < .001$. The mean number of units in essays of students in the Q&A condition increased from 1.7047 (*SD* = 1.54) to 4.34 (*SD* = 2.16). For essays of students in the Text condition this mean increased from 1.88 1.75 (*SD* = 1.25) to 4.86 (*SD* = 2.29).

Results are similar when we consider only evidence-based idea units. A GLMM showed a difference over time, $F(3, 172) = 38.818$, $p < .001$. The interaction between group and time was not statistically significant, $F(1, 172) = .015$, $p = .903$. The fixed effect of time was statistically significant in the overall model, $F(1, 172) = 113.598$, $p < .001$. The fixed effect of group was not statistically significant, $F(1, 172) = .763$, $p = .384$. The Q&A group showed an increase in evidence-based segments – from 1.30 (*SD* = 1.11) to 4.28 (*SD* = 2.26), and the Text group from 1.51 (*SD* = 1.39) to 4.57 (*SD* = 2.44).

We turn next to what is our key outcome variable, functional evidence-based idea units. Given the focus of our investigation was the use of evidence in argument, only segments that included a claim and evidence (evidence-based units) were examined. Furthermore, as noted above, we include only those evidence-based units in which the writer makes a functional connection between evidence and claim, using the former in support of the latter. Whether the two groups following the intervention showed differences on this variable is thus our key research question. An analysis of evidence-based idea units successfully used functionally to support or weaken a claim showed an interaction between group and time. Although students in both groups became successful in making functional use of the evidence available to them, a greater improvement was observed in the Q&A

condition. A GLMM showed a difference between groups over time in the functional use of evidence, $F(3, 172) = 23.998$, $p < .001$. The interaction between group and time was statistically significant, $F(1, 172) = 5.128$, $p = .025$. The fixed effect of Time was significant, $F(1, 172) = 64.256$, $p < .001$, although the fixed effect of group was not significant, $F(1, 172) = .185$. Although students in both conditions improved from initial to final assessment, students in the Q&A condition showed greater improvement, from .95 (*SD* = 1.09) to 3.79 (*SD* = 2.45), than students in the text condition, from 1.45 (*SD* = 1.43) to 3.04 (*SD* = 2.04).

This picture becomes clearer when we go on to examine gains for the functional evidence-based subtypes, in particular the Weaken-other subtype. The use of evidence to weaken claims has been shown in previous work to be more challenging than use of evidence to support a claim (Hemberger et al., 2017).

Weaken-other usage. In the Q&A group, the proportion of students ever showing the weaken-other type increased from 12.8% to 70.7%, $p < .001$. In the Text group this proportion increased from 21.4% to 47.7%. The proportions at final assessment were significantly higher in the Q&A group than in the Text group, $X^2 (1, 85) = 4.637$, $p = .047$. A GLMM showed a difference between groups in weaken-other usage, $F(3, 172) = 17.955$, $p < .001$. The interaction between group and time was significant, $F(1, 172) = 6.708$, $p = .010$. The fixed effect of time was significant, $F(1, 172) = 48.117$, $p < .001$, but the fixed effect of group did not reach significance, $F(1, 172) = 2.715$, $p = .101$. Students in the Q&A condition showed an increase in the number of Weaken-other units, from .17 (*SD* = .39) to 1.41 (*SD* = 1.21), while students in the Text condition showed a more limited increase, from .29 (*SD* = .61) to .86 (*SD* = .99).

Support-own usage. For the less challenging and overall most frequently used type, in contrast, the Support-own type, both groups showed comparable mastery. A GLMM showed that the overall model was significant, $F(3, 172) = 10.806$, $p < .001$. The interaction between group and time was not statistically significant, $F(1, 172) = .825$, $p = .365$, nor was the fixed effect of condition, $F(1, 172) = .172$, $p = .679$. The fixed effect of time was statistically significant, $F(1, 172) = 28.867$, $p < .001$. Both the Text condition students and the Q&A students increased in number of M+ units from 1.14 ($SD = 1.19$) to 2.12 ($SD = 1.79$) and from .85 ($SD = .92$) to 2.23 ($SD = 1.76$), respectively.

Weaken-own and Support-other usage. Weaken-own and support-other statements were much less frequent overall, as would be expected given they were inconsistent with the student's position, and only a time effect appeared. A GLMM using the Poisson probability distribution showed no differences between groups in weaken-own usage, $F(3, 172) = 1.798$, $p = .149$. The interaction between group and time was not statistically significant, $F(1, 172) = .012$, $p = .911$. The fixed effect of group was not statistically significant in the overall GLMM model, $F(1, 172) = .036$, $p = .849$. The fixed effect of time was statistically significant, $F(1, 172) = 5.358$, $p = .022$. At initial assessment, there was almost no use of weaken-own by the Q&A group ($M = .04$, $SD = .16$) or the Text group ($M = .04$, $SD = .16$). At the final assessment there was a small increase in the weaken-own usage in both the Q&A and text conditions ($M = .12$, $SD = .33$ and $M = .14$, $SD = .34$, respectively).

Support-other units were similarly rare. None of the students included them at initial assessment and at the final assessment only two students in the Q&A group and one in the Text group ever included them. A Poisson regression showed that there was no statically significant difference between groups in the gains they exhibited in the support-other usage

from initial to final assessment (Wald chi-square test = .410; *df* = 1, *p* = .522). Students in both conditions showed no support-other units at initial assessment, and very few at the final assessment, *M* = .02 (*SD* = .15) for the Text condition, and *M* = .05 (*SD* = .22) for the Q&A condition.

1.3. Discussion

These results serve the objective of providing evidence of the superiority of the Q&A method over a traditional whole-text method in making available to students knowledge they need to engage in meaningful discourse about a topic they have little prior knowledge of. Over four pieces of evidence on average made their way into the essays of students in both the text and Q&A conditions. However, students in the Q&A condition were more successful in making functional use of this evidence than were students in the text condition, as reflected in a significant condition difference in functional units, particularly in the more challenging function of weakening. No liabilities of the Q&A method were observed relative to the text condition.

Pinpointing the mechanisms involved in the success of the Q&A method will require more investigation. We believe the information choice and especially the question-and-answer format, anticipating the potential use of the information, are key, but we cannot rule out the contribution of factors identified in the extensive literatures on reading comprehension and on cognitive load. Future research is also indicated to identify which of the components in the Q&A condition were key: To what extent did (a) the potential functional purpose the Q&A format provided, (b) the segmentation of the text in smaller pieces of information, and (c) their distribution over time, facilitate functional use of

evidence in an argumentive essay? These become questions worth asking given the overall facilitative effect of their presence in combination.

As part of Study 2, we will compare gains in argument skills in the present case of unfamiliar, knowledge-dense topics to gains observed in earlier studies over longer periods of engagement with more than one topic. Evidence use gains from initial to final essay in the present Study 1 show that students acquired at least some of the knowledge contained in the evidence statements made available to them, consistent with Zohar and Nemet's (2002) highly-cited report that knowledge gain is enhanced when introduced in the context of a problematized issue to be addressed. Yet, students in the present study in their final essays made reference to only a small portion of the information potentially available (about 4 of 24 pieces). We would not expect a student to incorporate all of it in an argumentive essay, and no doubt students had acquired more knowledge than made its way into the essay. As noted, they did include a variety of different pieces of information, giving us reason to believe that the provided information was in large part comprehensible and accessible to them. Still, an important concern is that students most often only made use of those pieces of information that could serve as evidence consistent with their own position. Evidence favoring the opposing position or damaging to their own was largely ignored, whether students were in the Q&A or Text conditions. This is a particular concern given a primary rationale underlying the curriculum is to support students in overcoming "myside bias" by attending to the opposing position and thoroughly examining its strengths and weaknesses, as well as the strengths and weaknesses of their own position (Kuhn et al., 2016a). We explore this concern further in turning to Study 2, which focuses more directly on knowledge gains.

2. Study 2

In Study 2 we turn to a social topic, again one unfamiliar to middle school students and requiring knowledge acquisition to enable meaningful discourse. Our objective is to investigate the extent and nature of knowledge gains, in order to establish the feasibility of achieving both skill and knowledge gains within the same intervention.

2.1. Method

2.1.1. Participants

Participants in Study 2 were middle-school students (grades 6-8; ages 11-14) attending a one-week summer program at a local university in a large urban area in the US. The program was one advertised as an introduction to debate and was offered free-of-charge to students in the surrounding neighborhood. Participants consisted of all students who signed up for the program and hence constituted the sample available to us. Most were from low-income families and attended local public schools. About two thirds were of Hispanic ethnicity, with the remaining African American, Asian, or Caucasian. Participants reported having either no or at most one prior experience with debate. (A minority having more experience attended a different section of the program not reported on here.) On a brief introductory informational questionnaire, a majority indicated no particular interest in debate and stated that they were attending in accord with parents' wishes.

In the first of three equivalent groups, randomly formed by the program's administrative staff, 18 students participated. The second and third groups contained 19 and 21 students respectively. The three groups were similar in age and of roughly equal gender distribution. The topic and instructional program were identical across groups, except for several minor variations to be described, made possible by the fact that the three groups

attended the program sequentially, during different weeks rather than all during the same week. All participants attended only one week.

Also participating in Study 2 were 49 middle-school students who constituted a comparison group. They were of the same age and grade range and from similar neighborhoods as the intervention group, and also all from the same school attended by many of the intervention group.

2.1.2. Procedure

The intervention group's activity consisted of the intervention described here, similar in major features to the Study 1 intervention, except for a denser participation schedule. Students attended for two consecutive days from 9am to 3pm. Each day included approximately three hours of instructional time divided into three 50-60-min sessions (with rest and snack breaks interspersed, preceded by a provided breakfast and followed by a provided lunch). For the remainder of the week, participants proceeded to a more traditional debate program not staffed by the present researchers, which they were told the first two days would prepare them for.

At the beginning of the first day, following a general introduction to the program, the following introduction to the topic was presented in writing and read aloud:

> One tax American citizens pay is social security tax. It funds the social security program that president Roosevelt began in 1935. Workers and employers each contribute a percentage of the worker's pay to fund the program. After workers reach about age 67, it provides them a monthly allowance for the rest of their lives. Some workers are unhappy about how

much the tax takes out of their paycheck. Also, there are worries the fund will run out of money because there are now many more older people.

Should the social security program continue in its present form? Or should people be able to save for their own retirement, perhaps with an emergency fund for old people who run out of money.

Participants were then asked to choose in writing which position they favored:

Workers must contribute some their pay to the government Social Security system.

Or

Workers can instead save for their own old age.

On this basis the group was divided into the *SS* and the *SAVE* teams, based on their choice, with the few undecided students assigned to a team, after discussion with them, so as to achieve teams of approximately equal size. Teams then divided into smaller groups of 3 or 4 to begin the first activity.

The procedure differed in only minor ways from that described for Study 1. In same-side small groups, students generated reasons supporting the group's position and assembled and evaluated a set of "Reason cards" representing them. Same-side pairs then engaged in dialogs with a rotating sequence of six opposing-side pairs. While waiting for the opposing pair to respond, the pair worked on the reflection sheets described in Study 1.

Students then returned to same-side small groups to prepare for a final "Showdown" whole-class debate. In the Showdown, one student at a time verbally debated someone from the opposing side. Whenever they wished, either side could call a one-minute "huddle" to

confer. Following a break, students wrote individual final position essays. In a final debrief activity, students were guided through a transcription of portions of the Showdown debate. Points were awarded for effective argumentive moves and a winning team declared.

The participation of at least two and sometimes three of the authors throughout the activities enabled us to assure that implementation occurred as intended. In addition, a colleague not involved in the work reviewed video and audio excerpts and confirmed that the method in all conditions was consistent with the implementation of the method as intended and described here.

The set of 27 pieces of evidence in Q&A format were made available to students in a manner similar to that of the Study 1 experimental condition[3]. The set appears in online Appendix A. The initial introductory set of three Q&A pieces was presented at the beginning of the first activity, with subsequent sets of three introduced at intervals thereafter, each set remaining available thereafter. Day 1 ended with the first half of the dialogs and Day 2 began with the second half. Participants were offered the opportunity to pose any additional questions of their own at any time and later receive answers, but few did so, in part because the rapid pace left little available time.

2.1.2.1. Design of variations in evidence presentation

The existence of the three separate groups described enabled us to introduce minor procedural variations across the three groups having to do with presentation of the evidence Q&As, for the purpose of better understanding the role of specific components, in order to see how these affect subsequent use of this evidence. In one group, the standard condition, a verbal prompt was included when the evidence pieces were made available: “Try to use this information in your arguments” (following the procedure used by Hemberger et al., 2017).

For the purpose of assessing whether this prompt was essential to students' use of the evidence, in a second group this prompt was subtracted and students were told simply: "Here's some information about the topic." These prompts were delivered with each presentation and prior to the final essay. As we will report here, this subtraction proved not to affect performance. We were concerned, however, by the finding we will report (and also noted in Study 1) that students most often ignored evidence not favorable to their own positions. In the third group, therefore, we added a prompt, to promote attention to incongruent (Support-other or Weaken-own) evidence: "Not all of the evidence is going to support your side; if it doesn't, see if you can deal with it." This prompt was delivered only at the first evidence presentation (to avoid seeming repetitive) and again prior to the essay. Finally, within the third group, we also manipulated the order of evidence presentation, to determine whether the facilitative order used in earlier work (Hemberger et al., 2017), an order of previously observed difficulty (evidence supportive of own position, then, weakening opposing position, and finally supportive of opposing or weakening of own position) – could be subtracted without negative effect. For half of the group, the evidence was presented in this presumed facilitative sequence. The remaining half were presented the identical set of evidence but in a random order.

2.1.2.2. Comparison group

As a basis for comparison to knowledge acquisition in the intervention group, a questionnaire was administered to the comparison group of 49 middle-school students. The questionnaire was administered in the classroom by the classroom teacher to two social studies classes. It took the form of an 8.5x11" sheet divided into four quadrants. At the top of the sheet appeared this information: "Social security (SS) is a program in which workers pay

a percentage of their pay each month. When they are old and retire, the government then sends them a monthly allowance." In the four quadrants of the sheet, students were asked to "list all of the things you know" about each of these four: Beneficial things about Social security, Beneficial things about Saving on your own, Negative things about Social security, and Negative things about Saving on your own (each of the quadrants had been appropriately labeled).

We chose this method of comparison, instead of a pretest-posttest assessment of intervention students' knowledge, to avoid the possibility of a pretest assessment priming intervention students' attention to acquiring such knowledge, which would not have occurred in the absence of such an assessment. In employing a posttest group comparison, we wished to give the comparison group every possible advantage in displaying the full extent of their knowledge. Intervention students were credited with topic knowledge only if they included it in their final essays. However, they had been very recently deeply engaged with the topic over two days and were motivated to make use of knowledge they had gained in order to make a strong argument favoring their preferred alternative over the opposing one. By not asking the comparison group students to meet the challenges of composing an argumentive essay, we gave the comparison group a decided advantage in revealing the extent of their knowledge. In addition, we identified for them the four relevant categories within which they could organize this knowledge (thus reducing the likelihood of their ignoring their non-preferred alternative or ignoring negative attributes of either alternative), a further advantage favoring the comparison group over the intervention group.

2.2. Results

As in Study 1, the data we report on come from participants' final individual essays,

but focusing now on the knowledge gained about the topic, a topic students of this age had negligible preexisting knowledge about. We began by applying the coding scheme used in Study 1, which serves to confirm that the intervention, administered in a more dense time frame than in previous studies, had nonetheless produced outcomes comparable to earlier ones.

2.2.1. Argument skills reflected in individual essays

As in Study 1, essays were divided into idea units consisting of a claim and supporting justification, and these units were then coded as serving one of the four functions: *Support own* (M+), *Weaken other* (O–), *Support other* (O+), and *Weaken own* (M–). (A fifth category, non-functional, was included for idea units not serving any discernible argumentive function.) Table 1 presents the mean number of units in final essays that were classified as serving each of the four functions. These results proved consistent with earlier studies.

Of the within- and across-group manipulations, two showed no effect. Performance of the groups prompted vs. not prompted to use the evidence were very similar. Hence, for simplicity, results for these two groups are combined. The manipulation within the third group (sequenced vs. random evidence presentation) also showed negligible effect and hence this group is reported on as a whole. The results for the first two groups are presented as a whole to ease comprehension of our major findings, as are results for the third group. All results were examined for outliers and none were detected. Breakdowns between the first two groups and for the subgroups of the third group appear in online Appendix B for readers wishing to see them.

The third group overall, however, showed superior performance to the other groups; their essays contained a mean of 6.76 functional idea units, compared to 3.98 for the other

two groups combined. Groups nonetheless showed similar differences across types of functional idea unit. Consistent with Study 1 and earlier studies (Hemberger et al., 2017), the majority of all students included M+ units in their essays, as well as O- units. Much more challenging was the inclusion of O+ units, which were discrepant with the student's own position; M- units, also discrepant, were even less prevalent. Because of the overall limited use of O+ and M- types, it is relevant to ask how many students *ever* included them in their essays. These proportions appear in Table 2.

Table 1

Mean Number of Evidence-based Functional Units of Each Type in Final Essays

	M+	O-	O+	M-
Incongruent-Evidence-Prompt Group	3.14	2.48	.76	.38
Remaining Intervention Groups	1.84	1.81	.30	.03

Note. N = 21 for Incongruent-Evidence-Prompt Group and 37 for Remaining Groups. *Support own* = M+, *Weaken other* = O–, *Support other* = O+, and *Weaken own* = M–.

Table 2

Percentages of Student Essays Showing Each Type of Evidence-based Functional Unit at Least Once in Final Essays

	M+	O-	O+	M-
Incongruent-Evidence-Prompt Group	81%	81%	43%	29%
Remaining Intervention Groups	84%	70%	19%	03%

As Tables 1 and 2 show, the Incongruent-Evidence-Prompt Group appears to do better, in both percentage of students ever including discrepant O+ or M- functional evidence-based statements in their essays and in the overall frequencies with which they do so. Nonetheless, it remains a minority of this group who include any, despite the abundant availability of appropriate evidence by the time students wrote their final essays.

A number of differences preclude direct comparison with Study 1 findings. Although the Study 2 comparison group included on average about the same number of functional evidence-based claims in their essays (about four) as did Study 1 participants, students did not have the same amount of information to work with in the two studies. This is especially so as students in Study 2 introduced some evidence drawn from their own knowledge and sometimes shared it with their peers, a phenomenon that did not occur in Study 1.

A more equivalent comparison is thus the percentages of participants who *ever* include functional evidence-based claims of the various subtypes, apart from the question of frequency. In both studies, a key concern is the functional inclusion of evidence addressing the opposing position and, especially, evidence supportive of that position. Enough participants in both studies showed one or both of these at least once to make comparisons meaningful. These percentages appear in Table 3 for Studies 1 and 2 participants, and, for

comparison, parallel percentages from earlier studies of longer duration.[4] (Percentages for Weaken-Own statements are omitted because of their relative infrequency across all studies.)

Table 3

Cross-study Comparison of Percentages of Participants Showing Weaken-Other or Support-Other Functional Evidence-based Statements in their Essays

	Weaken-Other (O-)	Support-Other (O+)
Study 2, Incongruent-Evidence-Prompt Group	81%	43%
Study 2, Remaining Groups	70%	19%
Study 1, Q&A Group	71%	05%
One-year intervention, final studied topic (Hemberger et al., 2017)	74%	37%
One-year intervention, unstudied new topic (Hemberger et al., 2017)	11%	0%
Two-year intervention, unstudied new topic (Kuhn & Moore, 2015)	04%	0%

Note. Effect sizes for the above studies, where available, were in a range from .196 to .590.

2.2.2. Knowledge acquisition

We turn now to a different, content-, rather than function-based, descriptive analysis to address the key question posed in Study 2, the knowledge students came to have regarding

the topic by the time of the final essay devoted to it. We chose to focus analysis on the Incongruent-Evidence-Prompt Group as the group who appeared to perform most favorably in terms of argumentive skill and therefore presumably benefited the most from the intervention. Presented in Table 4 is a list of every knowledge claim related to the social security topic that appeared in the essays of this group. These are organized in terms of the typology of four possible argumentive functions employed earlier, but in this case for the group as a whole and hence without reference to the writer's position, thus Supporting Side A, Weakening Side A, Supporting Side B, and Weakening Side B.

The term 'information' might be considered preferable to the term 'evidence' in referring to the knowledge claims in this list since inclusion in the list did not require that this information be cited in the service of an argumentive claim, although it typically was. Note also that not all such pieces of information made available to students appear in Table 4, while others do appear in Table 4 that were not made available as part of the intervention. The latter we can presume arose from a participant's own knowledge base or from that of a peer. We can assume such social transmission occurred during the intervention given the frequency of appearance of some of these in multiple students' essays, e.g., "The government might steal your money." We need not make assumptions regarding the factual correctness of these student-generated claims. We included them as long as they were plausible ones from the perspective of young people in this age group.

Students' joint engagement with the topic thus led to common knowledge gains, as would be anticipated in a collaborative learning context. Students of course showed variation in the language they used in making these knowledge claims. In Table 4, we represent each one in the simplest possible language that captures the claim and is closest to the actual

language the majority of students used. Also included for each claim is the number of students (of the 21 in this group) whose essays included it. Two of the authors achieved 100% agreement in identifying these knowledge claims.

Table 4

Knowledge Claims in Student Essays on Social Security (SS) Topic

Claims Favorable to Social Security (SS)

1. You are assured money for retirement.* (n=16)
2. SS is an easy way to save.* (n=8)
3. Your money is safe.* (n=3)
4. You won't run out of money. (n=7)
5. SS only requires 6.2% per month. (n=5)
6. Employers don't always provide pensions. (n=2)
7. SS helps people who earn little. (n=1)
8. You can still also save on your own. (n=1)
9. Your children won't have to support you. (n=1)
10. Most countries provide some kind of SS. (n=1)

Claims Unfavorable to Social Security (SS)

1. The government could run out of money for SS. (n=9)
2. The government might steal your SS money.* (n=10)
3. Your SS payments go to others, not you. (n=4)
4. You might not live long enough to get SS benefits. (n=3)
5. SS benefits aren't large enough.* (n=3)

6. Some people contribute more than others (n=1)

Claims Favorable to Individual Saving

1. You can set up an individual retirement account. (n=3)
2. You know where your money is.* (n=2)
3. Your money is available for emergencies.* (n=3)
4. Your money is available whenever you want it.* (n=2)
5. You can decide whether and how much to save. (n=2)

Claims Unfavorable to Individual Saving

1. You don't know how much you will need. (n=4)
2. You may not save enough and run out.* (n=6)
3. Saving is difficult and stressful.* (n=8)
4. Your savings may be lost or stolen.* (n=6)
5. If your money is available, you may spend it. (n=3)
6. Your children may not help support you in old age. (n=3)
7. People don't know enough about financial matters. (n=1)

Note. n indicates number of students who included this statement in the essay.

* indicates claim was also made by one or more students in the non-intervention comparison group.

Turning now to the central question of the knowledge gains students showed, the question was addressed by comparing the knowledge claims that appeared in the final essays of the intervention group to those appearing in a knowledge survey completed by a non-intervention group, a comparison that methodologically privileged the non-intervention

group in a number of respects, as described earlier in introducing the study design (Section 2.1.2.2). Considerable variation appeared in how many different claims appeared in students' essays. The number ranges from 2 to 11, with a median of 6 and a mean of 6.05. (Scores did not differ significantly as a function of the manipulation within this group – sequenced vs. random presentation order.)

The 49 students in the comparison group identified a total of 47 knowledge claims – slightly less than one per student (compared to the slightly more than six per student in the intervention group 3, as reported). Modal performance of the comparison group was zero (the score of 23 of the 49 students), with a range from zero to four. Comparison group knowledge claims duplicated some of those of the intervention group (indicated by an asterisk in Table 4), supporting the conclusion that not all of the intervention group's knowledge was entirely new to their participation in the intervention. Most common were "SS is an easy way to save" (listed by nine of 49 comparison students) and on the Save side "You may not save enough" (listed by seven of 49). Others were expressed by five or fewer students[5]. Other claims made by comparison group students, however, not seen in the intervention group, and not credited as valid knowledge claims, were change-worthy incorrect knowledge claims. A total of 14 appeared. On the social security side, for example, were these: "It pays for everything you need" and "It prevents people from pretending to be you." On the saving side, these were all variations of the same theme: "You'll have enough money," "You'll have a lot of money," "You'll have more money."

2.3. Discussion

The assessment of topic knowledge on the part of a comparison group gives us confidence in believing that significant knowledge acquisition took place on the part of the

intervention group. What does this assessment in the knowledge-rich context of Study 2 tell us about the nature and process of such knowledge acquisition? Comparing the evidence-based claims as participants expressed them in their essays and the evidence as it was presented to them during the intervention) shows that few evidence-based claims in the essays include precise reiterations of this evidence as it was initially encountered. Rather, the information presented is registered by the student but then integrated with pre-existing knowledge, serving to make it congruent with the student's thinking more broadly, as proposed in the introduction to Study 1. A striking example is the information that social security tax is collected regularly by the government from an employee's earnings. Once acquired, students used this knowledge to make the claim that this money would be kept safe; others, however, used the same information as a basis for the claim that the money could be stolen.

Of particular interest is how students integrated knowledge that was not congruent with their position and could potentially weaken it. This question is both theoretically and pragmatically important, given students most frequently fail to address such evidence, as noted earlier. The procedural difference between Group 3 and the other two intervention groups was a seemingly slight one. ("Not all of the evidence is going to support your side; if it doesn't, see if you can deal with it," they were told once earlier and once just before they wrote their essays, whereas Group 1 was told only to try to make use of the information and Group 2 only that is was available.) Yet Group 3 showed greatest success in acknowledging incongruent evidence (43% vs. 19%, Table 3). Thus, even seemingly minor changes in instructional framing (in this case, from "try to use" to "try to deal with" the discrepant evidence) can make significant differences in how students construe their task. The "try to

deal with" language we believe was more effective because it acknowledges the incongruity by signaling to the student that the statement is not going to serve a supportive function. Nonetheless, despite the notable effect of this minimal prompt, further investigation is needed to establish the extent to which the gains observed were highly context-tied to this explicit prompt or would be maintained across time and generalize to diverse contexts.

How did the almost half of Group 3 students who were able to "deal with" evidence incongruent with their positions do so? Some simply acknowledged it but left it unaddressed and turned to other evidence that did support their position. For example, a student who favored the social security position stated in his essay:

> People who don't support social security say the government steals your money, and will run out of money. I actually think that social security is good because if you save you don't know if you will save enough.

Others acknowledged the incongruent evidence but then produced a counterargument against it, as in these two examples:

> You think that your kids can give you money. But what if you don't have kids, what if your kids hate you, what if your kid doesn't have the money you need.

> You're probably saying what if you have an emergency? Social security only requires 6.2 percent of your paycheck.

The remaining half of students, as we saw, in their essays simply neglected to mention evidence that did not support their position. This happened even when the student had addressed this evidence and sometimes even successfully countered it during the preceding dialogs with their opposing-side peers. For example, in the course of their dialogs one pair in favor of social security twice made the claim that the government keeps your money safe for you, in dialogs with two different opposing pairs. Each time their opponents on the save side countered this claim, claiming that contributions went to current recipients rather than being held for the contributors themselves. During the second of these dialogs, the social security pair conceded the point. In their final essays, however, one of the social security pair ignored the issue, while the other resurrected the pair's original claim that your contributions are kept for you, omitting any mention of the counterargument he had at that point heard twice and acknowledged. More positive regarding knowledge acquisition, however, are observations of progress in mastery of even more complex conceptual knowledge that goes beyond the simple factual. One essay states, for example, "The middle and lower class get paid a smaller amount, so that 6.2% really does hurt their check."

3. Conclusion

Studies 1 and 2 fulfill the objective of establishing that it is possible to accomplish both knowledge and skill goals in the context of a curriculum centered around dense engagement in argumentation. Furthermore, it is possible to do so even in those cases in which pre-existing factual knowledge regarding the topic is non-existent or at most minimal and superficial, as was the case for the topics involved in both the studies presented here. Establishing the feasibility of achieving these dual goals in a single instructional context is a

necessary first step in closer examination of the interface between progress on the knowledge front and progress on the skill front, an empirical question that deserves further investigation (Schwarz & Baker, 2016). The illustrations presented in connection with Study 2 provide some ideas regarding how newly received knowledge is absorbed, transformed, and gradually put to use as mastery of it becomes firmer and skill level allows. In any case it is clear that knowledge gain is not limited to cases of restructuring of existing knowledge, as is sometimes the case in studies of conceptual change.

Enough knowledge must be acquired, however, to make argumentation productive or even feasible. The Q&A method we examined in Study 1 for accomplishing this goal is arguably more than one of breaking information into more manageable bits. A major strength of this method, we have claimed, is in enabling students to recognize the potential purpose and therefore value of such information.

Taken together, Studies 1 and 2 provide further empirical support for treating discourse as a productive path to individual written argument and in particular for a method that emphasizes direct peer-to-peer discourse (Kuhn et al., 2008; Kuhn et al., 2016a,b; Kuhn, 2018). Given the difficulty that students of all ages have with non-fiction writing, this finding is of broad practical significance, especially to the degree it extends to technical, complex subject matter students have limited familiarity with, as was the case in the present study.

Study 2 also documents that the dense all-day "workshop" mode of participation in the curriculum, not reported on previously, is effective for students of this age group. Indeed, our qualitative impression is that this mode is if anything more effective than one of lesser density, given that the sequential nature of the activities builds to a climax and students

become increasingly absorbed, in contrast to their typical school experience of shifting frequently from subject to subject throughout the school day.

Among the limitations of the present work is its confinement to two highly specific, knowledge-dense topics that students the age of the study's participants had minimal knowledge of. The studies are thus limited by their "situativeness" with respect to both content and context. Would the present results be replicated in different kinds of settings and/or with topics that participants of this age already have significant knowledge and strong opinions about? These are questions for future research.

The present findings suggest that it is not the existence of previous knowledge about the topic that is the key to successful student writing. Familiar topics are perhaps where novice writers should start (Calkins, 1994), but topics need not stop there – a lack of knowledge can be remedied, as the present studies confirm**.** Instead, the other key to the success reported in the present studies in our view is the deep immersion in a topic. As reflected in Table 3, skills are much more modest when writers move to a new, unstudied topic.

A further benefit of dialogic argumentation is its fostering of what Nussbaum and Asterhan (2016) refer to as *proactive executive control strategies*. These are as fully necessary in writing as they are in discourse. In writing, there is no external other to serve this executive control function and to instill purpose (Graff, 2003). The discourse skills students develop in peer dialog are of course valuable in and of themselves, not simply as a bridge to effective argumentive writing. Similarly, the knowledge students gain in arguing to learn is important in its own right, but its benefits do not end there. Arguing to learn develops the rigorous thinking that underlies accomplished argumentive writing. As an

individual or social practice, it enables us to clarify the basis for our commitments in a framework of alternatives and evidence. Attending to both knowledge and skill goals, then, as we sought to do in the present work, can be seen as an ideal.

Footnotes

1. Some examples of the information included in the Q&A cards in Study 1 are the following: “How was natural gas created? Natural gas was created millions of years ago in the depths of the oceans after animals and plants died and were covered by large mud and sand layers. There, high pressure and heat converted the organic material into oil and gas. Natural gas is trapped in underground geological cavities. Natural gas is considered a non-renewable source of energy because it takes millions of years to be created.” and “How can solar energy be used? The first direct use of solar energy is the heating of the people themselves and, of the spaces that they live and work in and the water they use. The heat provided by the sun is also utilized in agriculture greenhouses and for drying products. There are three ways to use solar energy. (1) Utilizing the warmth of the sun for heating, cooling and heating water with solar thermal systems. (Active solar systems), (2) constructing bioclimatic buildings (passive solar) and (3) producing electricity through photovoltaic systems.”
2. GLMM was pursued in one imputed dataset. The results were similar when other imputed datasets were used. Details are available from authors.
3. For example, the following pieces of evidence were provided to students in Study 2 in the form of Q&A “How much must American workers and employers pay to the Social Security fund? In the US, workers pay 6.2% of every paycheck and the employer pays another 6.2%. In some countries, however, the required contribution is much higher. One of the highest is Italy, where the employer contributes 33% of every paycheck and the worker contributes 9%. In a few countries, such as Armenia, only the employee contributes, not the employer” and “How many people over 65 live in poverty? This varies by country. In the US, 20% have incomes less than half of the average person.”

4. Percentages are higher in studies in which the criterion for crediting an idea unit is simply making a reasoned claim (Kuhn & Crowell, 2011; Kuhn et al., 2008) rather than being functionally evidence-based.

5. Two additional valid claims unfavorable to social security that had not appeared in the intervention group appeared among the comparison group, expressed by one participant each: “You won’t have the money till you’re old” and “If you owe money, paying your SS tax could leave you in debt.” Not given credit as a valid knowledge claim were statements that did not distinguish the opposing alternatives or offer a justification for either alternative (e.g., “You can earn money by having a job” or “Some people save for a special occasion”) or simply repeated verbatim the given information (“The government sends them a monthly allowance”).

References

Andriessen, J., & Baker, M. (2014). Arguing to Learn. In R. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (Cambridge Handbooks in Psychology, pp. 439-460). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139519526.027

Asterhan, C., & Schwarz, B. (2016). Argumentation for learning: Well-trodden paths and unexplored territories. *Educational Psychologist, 51,* 164-187.

Barzilai, S., & Chinn, C. A. (2018). On the goals of epistemic education: Promoting apt epistemic performance. *Journal of the Learning Sciences, 27*(3), 353-389.

Bereiter, C., & Scardamalia, M. Fixing Humpty Dumpty: Putting higher-order skills and knowledge together again. In L. Kerslake & R. Wegerif (Eds.) (2018). *Theory of teaching thinking: International perspective.* London: Routledge.

Calkins, L. (1994). *The Art of Teaching Writing.* New York: Heinemann.

Creswell, J. W. (2014). *Educational Research: Planning, Conducting and Evaluating Quantitative and Qualitative Research*. Harlow: Pearson.

Dewey, J. (1916). Democracy and education: An introduction to the philosophy of education. In J. Boydston (Ed.), *The Middle Works of John Dewey* (Vol. 9). Carbondale, IL: Southern Illinois University Press.

Enders, C. K. (2010). *Applied missing data analysis*. Guilford press.

Elmore, R. (2018). Beijing Academy: Innovation, design, and learning. *ECNU Review of Education, 1,* 135-146.

Graff, G. (2003). *Clueless in Academe: How Schooling Obscures the Life of the Mind.* New Haven: Yale University Press.

Hemberger, L., Kuhn, D., Matos, F., & Shi, Y. (2017). A dialogic path to evidence-based argumentive writing. *Journal of the Learning Sciences, 26*, 575-607.

Hoaglin, D. C., & Iglewiez, B. (1987). Fine tuning some resistant rules for outlier labeling. *Journal of American Statistical Association, 82*, 1147-1149.

Iordanou, K. & Constantinou. C. P. (2015). Supporting use of evidence in argumentation through practice in argumentation and reflection in the context of SOCRATES learning environment. *Science Education, 99*, 282–311.

Kuhn, D. (2018). *Building our best future: Thinking critically about ourselves and our world*. New York: Wessex Learning.

Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. Psychological Science, 22, 545 – 552.

Kuhn, D., Hemberger, L., & Khait, V. (2016). Tracing the development of argumentive writing in a discourse-rich context. Written Communication, 33, 92-121. (a)

Kuhn, D., Hemberger, L., & Khait, V. (2016). Argue with me: Argument as a path to developing students' thinking and writing. New York: Routledge. (2nd ed.) (b)

Kuhn, D., & Moore, W. (2015). Argument as core curriculum. Learning: Research and Practice, 1, 66-78.

Murphy, P.K., Greene, J., Firetto, C., Hendrick, B., Li, M., Montalbano, C., & Wei, L. (2018). Quality talk: Developing students' discourse to promote high-level comprehension. *American Educational Research Journal,* Advance online publication. https://doi.org/10.3102/0002831218771303

Nussbaum, E.M., & Asterhan, C.S. (2016). The psychology of far transfer from classroom argumentation. In F. Paglieri (Ed.), *The Psychology of Argument: Cognitive Approaches to Argumentation and Persuasion*. London, U.K.: College Publications.

Resnick, L., Asterhan, C., & Clarke, S. (2015). *Socializing intelligence through academic talk and dialogue*. American Educational Research Association.

Schwarz, B. B., & Baker, M. J. (2016). *Dialogue, Argumentation and Education: History, Theory and Practice*. Cambridge University Press.

Van der Veen, C., & Van Oers, B. (2017). (Eds.), Classroom dialogue and learning outcomes. *Learning and Instruction, 48,* 1-70.

Von Aufschnaiter, C., Erduran, S., Osborne, J., & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific through dilemmas in human genetics. *Journal of Research in Science Teaching, 45,* 101-131.

Wecker, C., & Fischer, F. (2014). Where is the evidence? A meta-analysis on the role of argumentation for the acquisition of domain-specific knowledge in computer-supported collaborative learning. *Computers & Education, 75,* 218-228.

Wegerif, R. (2012). *Dialogic: Education for the Internet Age.* London: Routledge.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*, 71-95.

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching, 39,* 35-62.

Online Appendix A

The Questions and Answers that Were Provided to the Q&A Condition in Study 1

Natural Gas

What is natural gas?

Natural gas is an energy source. Natural gas is a gaseous mixture consisting mainly of methane. The composition of the gas differs depending on its source. It is colorless, amorphous and odourless in its pure form. The characteristic smell given artificially for safety purposes, so as to be able to detect a leakage.

How was natural gas created?

Natural gas was created millions of years ago in the depths of the oceans after animals and plants died and were covered by large mud and sand layers. There, high pressure and heat converted the organic material into oil and gas. Natural gas is trapped in underground geological cavities. Natural gas is considered a non-renewable source of energy because it takes millions of years to be created.

How can we get natural gas?

Detection of Gas is not easy. Imagine that you have a buried treasure for which you don't have a map leading to it, and that this treasure is also invisible. Geologists are using technology, such as artificial seismic waves, trying to predict where they can find the gas so as to mine it. The natural gas fields are usually away from the main consumption centers.

How can natural gas be transported?

The transportation of natural gas depends on its condition. In gaseous form, it is transported through pipelines at high pressure, and in liquid form is transported by tankers. The large high- pressure pipelines enable the transportation of gas for thousands of kilometers. An example of such pipelines is the pipeline which extends from Siberia to Central and Western Europe.

Which countries have natural gas?

Countries with large stocks of natural gas are Russia with 48,700,000,000,000 cubic meters, Iran with 33,600,000,000,000 cubic meters, Qatar with 24,700,000,000,000 cubic meters, Turkmenistan (Central Asia) with 17,500,000,000,000 cubic meters and the United States of America with 9,860,000,000,000 cubic meters.

Does Cyprus have natural gas?

In the sea area of Cyprus, there are ongoing investigations for the detection of Natural Gas. The Noble company that conducts surveys in the sea area of Cyprus and specifically in plot 12 supports that has identified a rich gas field, of 300 billion cubic meters of gas. This can bring to Cyprus earnings of half a billion euros annually.

Is natural gas good or bad for the environment?

Although the burning of natural gas does not produce sulfur compounds that pollute the environment and cause the phenomenon of acid rain, it does produce carbon dioxide. Carbon dioxide is produced during the mining of natural gas. For example 69.3 tons of carbon dioxide is produced every year from gas mining platforms in the United States.

How does natural gas compare to other fuels?

Burning natural gas produces lower amounts of carbon dioxide for each unit of energy compared with other fuels, and therefore has less harmful effects on the environment. Specifically, the combustion of natural gas produces 177 grams of carbon dioxide per kilowatt hour while burning oil produces 249 grams of carbon dioxide per kilowatt hour.

What kinds of bad effects can natural gas have?

Many of the areas which have being explored for natural gas are natural areas and the development of these areas can have negative effects on the natural environment and the wildlife in the area. There may be also toxic gas leaks that are detrimental to fishes. In addition to the above, in areas which have been explored for natural gas in the US, by drilling holes into their ground, more earthquakes have been observed.

What is the cost of producing electricity using natural gas?

Natural gas has reduced operational fuel management and maintenance costs. The cost for generating power using natural gas is between 0.07-0.13 euro per kWh.

Where can natural gas be used?

The natural gas has a high thermal efficiency, compared to that of other conventional fuels such as coal or oil. Natural gas before its use as a fuel must be treated so as to remove all the substances it contains besides methane. Natural gas is used in several ways and in different

fields, such as electricity, heating, heating water, cooking / baking, transportation (buses and cars) and industry.

What is the most desirable feature of natural gas?

The major characteristic of natural gas that make it a favorable option for the industry is the possibility for continuous supply of fuel, thus ensuring the smooth operation of industries.

Solar Energy

What is solar energy?

Solar Energy is the energy that comes from the sun and is considered a renewable energy source, since the sun is an endless source of energy.

How can solar energy produce electricity?

The exploitation of solar radiation and its conversion into electricity is now possible through Photovoltaic Technology. The solar radiation falls on the solar panels and it is transformed into electricity. The photovoltaic systems can be placed on land, roofs (flat and inclined) or on building facades.

How does solar energy work?

One can use photovoltaic systems in two ways; either independently of the Electricity Authority of Cyprus (EAC)'s network or connecting them with it. A photovoltaic (PV) system can therefore constitute an autonomous system that covers the entire energy needs of

a building or a business unit. In such a case, a storage unit (battery) needs to also be installed so that consumers will receive continuous electricity. Alternatively, a power generation system with solar panels can be used in conjunction with the EAC grid. In this case, one will consume power from the grid when the PV system's output is insufficient (e.g., a cloudy day or during the night) and gives energy to the grid when the output exceeds the needs, e.g., on very sunny days or when you are not at home.

Is solar energy good or bad for the environment?

The PV systems that produce solar energy operate without noise. They require very little maintenance and they do not pollute the environment. They have a long life and can meet the electricity needs in remote areas where there is no electricity network. The PV systems can be used as building materials, replacing other traditional materials (e.g., glazed roofs or facades). In this way they save money and natural resources.

Does Cyprus have solar energy?

Cyprus receives abundant sunlight. The solar energy that arrives via radiation every day in Cyprus can meet the energy needs of Cyprus' residents for half a year. Cyprus is a country with a remarkable potential for solar energy, since there is high radiation throughout the year, with average daily sunshine from 9.8 to 14.5 hours.

Does Cyprus have the technology required for using solar energy?

The technology of active solar systems for hot water production has wide commercial application in Cyprus, as 92% of households and 53% of hotels have solar water heating

systems, which according to a study of the European Union makes Cyprus a pioneer in the field of thermal solar energy applications.

How does Cyprus compare with other countries in use of solar energy?

Cyprus ranks first among the newly accepted 17 countries in the European Union in terms of installed PV (photovoltaic) systems. By April 2008 Cyprus had photovoltaic systems of total power 1.480 KW, of which 956,5 KW was connected to the electric network.

What kinds of bad effects can solar energy have?

Solar energy is environmentally clean, since it does not involve any polluting process. A typical solar power of 1 kilowatt (kW) produces an average of 1200-1500 kilowatt hours per year (depending on the sunlight of the region) and prevents on average 1,450 pounds of carbon dioxide which would have been released annually if alternative sources were used. This amount of carbon dioxide corresponds to the amount of carbon dioxide that can be absorbed by two acres of forest. However, the solar water heaters on the roofs of houses and apartment buildings pollute the visual environment of an area.

Is it expensive to produce electricity using solar power and can Cyprus afford it?

The power cost for solar water heating systems is 0.24 euro per kilowatt hour and for photovoltaic systems is 0.13 Euro per kWh. The photovoltaic systems have very high construction cost. In the situations where there is a need to store the electricity produced, there is the additional cost of the battery whose life is short, creating additional maintenance and replacement cost.

In winter, in the mountainous areas the water pipes freeze and break so solar water heating systems may need to be replaced. Yet, the high cost of PV systems can be depreciated in the long term by saving money on electricity bills, and selling the generated electricity of PV to the EAC network.

How can solar energy be used?

The first direct use of solar energy is the heating of the people themselves and, of the spaces that they live and work in and the water they use. The heat provided by the sun is also utilized in agriculture greenhouses and for drying products. There are three ways to use solar energy. (1) Utilizing the warmth of the sun for heating, cooling and heating water with solar thermal systems. (Active solar systems), (2) constructing bioclimatic buildings (passive solar) and (3) producing electricity through photovoltaic systems.

Is it feasible to use only solar energy to generate electricity?

Practically it is difficult to produce electicity only by solar energy. For example, in order to meet the needs of the USA in electricity, it requires 44,000 square kilometers, which corresponds to a size larger than Switzerland. Additionally, it would take 130.8 square kilometers for the storage of batteries. The cost of batteries is also too high, costing 3.3 trillion euros.

Is it feasible to construct photovoltaic systems to meet our needs in electricity?

Assuming that 1 meter squares photovoltaic panels is manufactured in 1 second, it would take 930 years to construct panels and 29.3 billion in order to meet the needs only of the USA in electricity.

The Texts Provided to the Text Condition in Study 1

Natural Gas

Natural gas is an energy source. Natural gas is a gaseous mixture consisting mainly of methane. The composition of the gas differs depending on its source. It is colorless, amorphous and odourless in its pure form. The characteristic smell given artificially for safety purposes, so as to be able to detect a leakage. Natural gas was created millions of years ago in the depths of the oceans after animals and plants died and were covered by large mud and sand layers. There, high pressure and heat converted the organic material into oil and gas. Natural gas is trapped in underground geological cavities. Natural gas is considered a non-renewable source of energy because it takes millions of years to be created.

Detection of Gas is not easy. Imagine that you have a buried treasure for which you don't have a map leading to it, and that this treasure is also invisible. Geologists are using technology, such as artificial seismic waves, trying to predict where they can find the gas so as to mine it. The natural gas fields are usually away from the main consumption centers. The transportation of natural gas depends on its condition. In gaseous form, it is transported through pipelines at high pressure, and in liquid form is transported by tankers. The large high- pressure pipelines enable the transportation of gas for thousands of kilometers. An example of such pipelines is the pipeline which extends from Siberia to Central and Western Europe.

Countries with large stocks of natural gas are Russia with 48,700,000,000,000 cubic meters, Iran with 33,600,000,000,000 cubic meters, Qatar with 24,700,000,000,000 cubic meters, Turkmenistan (Central Asia) with 17,500,000,000,000 cubic meters and the United States of America with 9,860,000,000,000 cubic meters. In the sea area of Cyprus, there are ongoing investigations for the detection of Natural Gas. The Noble company that conducts surveys in the sea area of Cyprus and specifically in plot 12 supports that has identified a rich gas field, of 300 billion cubic meters of gas. This can bring to Cyprus earnings of half a billion euros annually.

Although the burning of natural gas does not produce sulfur compounds that pollute the environment and cause the phenomenon of acid rain, it does produce carbon dioxide. Carbon dioxide is produced during the mining of natural gas. For example 69.3 tons of carbon dioxide is produced every year from gas mining platforms in the United States. Burning natural gas produces lower amounts of carbon dioxide for each unit of energy compared with other fuels, and therefore has less harmful effects on the environment. Specifically, the combustion of natural gas produces 177 grams of carbon dioxide per kilowatt hour while burning oil produces 249 grams of carbon dioxide per kilowatt hour.

Many of the areas which have being explored for natural gas are natural areas and the development of these areas can have negative effects on the natural environment and the wildlife in the area. There may be also toxic gas leaks that are detrimental to fishes. In addition to the above, in areas which have been explored for natural gas in the US, by drilling holes into their ground, more earthquakes have been observed.

Regarding cost, natural gas has reduced operational fuel management and maintenance costs. The cost for generating power using natural gas is between 0.07-0.13 euro per kWh.

The natural gas has a high thermal efficiency, compared to that of other conventional fuels such as coal or oil. Natural gas before its use as a fuel must be treated so as to remove all the substances it contains besides methane. Natural gas is used in several ways and in different fields, such as electricity, heating, heating water, cooking / baking, transportation (buses and cars) and industry. The major characteristic of natural gas that make it a favorable option for the industry is the possibility for continuous supply of fuel, thus ensuring the smooth operation of industries.

Solar Energy

Solar Energy is the energy that comes from the sun and is considered a renewable energy source, since the sun is an endless source of energy. The exploitation of solar radiation and its conversion into electricity is now possible through Photovoltaic Technology. The solar radiation falls on the solar panels and it is transformed into electricity. The photovoltaic systems can be placed on land, roofs (flat and inclined) or on building facades. They have a long life and can meet the electricity needs in remote areas where there is no electricity network. One can use photovoltaic systems in two ways; either independently of the Electricity Authority of Cyprus (EAC)'s network or connecting them with it. A photovoltaic (PV) system can therefore constitute an autonomous system that covers the entire energy needs of a building or a business unit. In such a case, a storage unit (battery) needs to also be installed so that consumers will receive continuous electricity. Alternatively, a power generation system with solar panels can be used in conjunction with the EAC grid. In this case, one will consume power from the grid when the PV system's output is insufficient

(e.g., a cloudy day or during the night) and gives energy to the grid when the output exceeds the needs, e.g., on very sunny days or when you are not at home.

The PV systems that produce solar energy operate without noise. They require very little maintenance and they do not pollute the environment. The PV systems can be used as building materials, replacing other traditional materials (e.g., glazed roofs or facades). In this way they save money and natural resources. A typical solar power of 1 kilowatt (kW) produces an average of 1200-1500 kilowatt hours per year (depending on the sunlight of the region) and prevents on average 1,450 pounds of carbon dioxide which would have been released annually if alternative sources were used. This amount of carbon dioxide corresponds to the amount of carbon dioxide that can be absorbed by two acres of forest. However, the solar water heaters on the roofs of houses and apartment buildings pollute the visual environment of an area.

Cyprus receives abundant sunlight. The solar energy that arrives via radiation every day in Cyprus can meet the energy needs of Cyprus' residents for half a year. Cyprus is a country with a remarkable potential for solar energy, since there is high radiation throughout the year, with average daily sunshine from 9.8 to 14.5 hours. The technology of active solar systems for hot water production has wide commercial application in Cyprus, as 92% of households and 53% of hotels have solar water heating systems, which according to a study of the European Union makes Cyprus a pioneer in the field of thermal solar energy applications. Cyprus ranks first among the newly accepted 17 countries in the European Union in terms of installed PV (photovoltaic) systems. By April 2008 Cyprus had photovoltaic systems of total power 1.480 KW.

The power cost for solar water heating systems is 0.24 euro per kilowatt hour and for photovoltaic systems is 0.13 Euro per kWh. The photovoltaic systems have very high construction cost. In the situations where there is a need to store the electricity produced, there is the additional cost of the battery whose life is short, creating additional maintenance and replacement cost. Another problem is that in the mountainous areas in winter, the water pipes freeze and break so solar water heating systems may need to be replaced. Yet, the high cost of PV systems can be depreciated in the long term by saving money on electricity bills, and selling the generated electricity of PV to the EAC network.

There are three ways to use solar energy. (1) Utilizing the warmth of the sun for heating, cooling and heating water with solar thermal systems. (Active solar systems), (2) constructing bioclimatic buildings (passive solar) and (3) producing electricity through photovoltaic systems. The heat provided by the sun is also utilized in agriculture greenhouses and for drying products.

Practically it is difficult to produce electricity only by solar energy. For example, in order to meet the needs of the USA in electricity, it requires 44,000 square kilometers, which corresponds to a size larger than Switzerland. Additionally, it would take 130.8 square kilometers for the storage of batteries. The cost of batteries is also too high, costing 3.3 trillion euros. Assuming that 1 meter squares photovoltaic panels is manufactured in 1 second, it would take 930 years to construct panels and 29.3 billion in order to meet the needs only of the USA in electricity.

Initial Social Security Evidence Questions and Answers provided in Study 2

Is the Social Security benefit paid to older people the same as a pension?

No. Pensions are paid by a specific employer to their own employees when they retire. Social Security is a government program that all employers and employees must contribute to.

How much must American workers and employers pay to the Social Security fund?

In the US, workers pay 6.2% of every paycheck and the employer pays another 6.2%. In some countries, however, the required contribution is much higher. One of the highest is Italy, where the employer contributes 33% of every paycheck and the worker contributes 9%. In a few countries, such as Armenia, only the employee contributes, not the employer.

Do all countries have a Social Security program?

No. Most countries provide some kinds of help to their older citizens, but in some countries (such as Turkey, Greece, and India) it is paid for at least partly by the government's general funds that come from many types of taxes. A few countries that have a Social Security program similar to the one in the US are Brazil, Germany, Kenya, Singapore, and Sweden.

Q&As Favorable to Social Security Program

SS+1. **Can an older person be sure they will get Social Security benefits?**

Yes, if they have contributed when they were working. Everyone who has contributed is entitled, no matter how much money they already have.

SS+2. **At what age can you begin to get Social Security benefits?**

The age varies by country and ranges from as young as 50 to about 65. In the US, you must be 67 to get full benefits. You can retire beginning at age 62 but your benefit will be lowered.

SS+3. **Does a person need to be retired to get Social Security benefits?**

No, in most countries if you have reached the required age, you can get benefits and continue to work. Some exceptions are Germany, Greece, and Spain, where you must be retired.

SS+4. **Do all older persons get the same Social Security benefit?**

No. The average benefit in 2015 in the US is $1230 per month, but some people get much less and some much more. In most countries it varies depending on how much you contributed while working. Generally, the more you contributed the more you receive in benefits (but in some countries, such as Armenia, it depends only on number of years you worked).

SS+5. **What happens if an older person has never worked or contributed to the Social Security system?**

Most countries, including the US, help such people, but they must show they are in need and they do not get as big a benefit as workers do.

SS+6. **If the Social Security program runs out of money, what will happen?**

a. Most likely, people will still get benefits, but changes will need to be made. Some possibilities are: Give benefits only to older people who need them and not to those who have enough money to take care of themselves.

b. Increase age at which people can start to receive benefits (now age 67 in US).

c. Increase percentage of their pay that workers and employers must contribute.

d. Use money from other government sources, such as sales tax.

Q&As Unfavorable to Social Security Program

SS-1. **Does the Social Security program cost the government a lot?**

Yes. In the US, 37% of government spending in 2013 was for Social Security benefits.

SS-2. **Is the money that is subtracted from workers' paychecks kept safe for them until their old age?**

No. The money is used for benefits to today's older people. When today's workers are old, new younger workers will contribute the money for benefits to today's workers.

SS-3. **Do all workers contribute the same percentage of their salary to the Social Security program?**

No, workers who earn more than $117,000 per year contribute only 6.2% of $117,000. So a worker who earns $117,000 per year and one who earns $517,000 per year contribute the same amount: 6.2% of $117,000.

SS-4. **If instead of contributing to the Social Security program workers had this money to use as they wanted to, would they be better off?**

This depends what they do with the money. Some claim they could invest the money themselves and would end up with a larger amount in their old age than they will receive from social security.

SS-5. **Is it possible for the US Social Security program to run out of money?**

Yes. It needs to bring in enough money from the contributions of working people to pay the older people who receive Social Security benefits. In countries where old age benefits come

from general government funds, this is less likely to happen, because the government can use money from other taxes, such as sales tax, to pay for the benefits to old people.

SS-6. **Is the US Social Security program likely to run out of money?**

There are different views, but most agree the system will not have enough money to keep paying the level of benefits old people receive today. One estimate is that this will happen in 2033.

Q&As Favorable to Individual Saving

SAV+1. **Do ways exist to help workers save?**

Yes. Individual Retirement Accounts (IRAs) are now widely available. An individual can choose to invest money in one and keep control of it, taking it out if needed or saving it for old age.

SAV+2. **Do employers give their employees a pension when they retire?**

Many private companies and all government offices worldwide provide their employees a pension when they retire that gives them a monthly income for the rest of their lives. This pension is separate from social security. The amount varies greatly.

SAV+3. **Can people be taught about managing their money?**

Classes exist in the US and other countries to teach financial skills and money management to children and adults. They are part of the curriculum in about half of US states. In parts of China all schools have them.

SAV+4. **Can people be helped to think more about their future selves in old age?**

One study says yes. After seeing computer-generated photos of themselves as older people, young people were more willing to save or invest, rather than spend, $1000 they were told they would receive,

SAV+5. **Must older people leave their homes if they run out of money?**

If people own their homes, reverse mortgages are available to help people stay in their homes. These allow home owners to receive a monthly payment from a lender. The lender will be paid back when the house is sold after the owner has died.

SAV+6. **Do grown children often help to support their parents?**

About one in five US middle-aged adults have provided financial support to a parent. In many countries the percentage who do is much higher.

Q&As Unfavorable to Individual Saving

SAV-1. **How many people over 65 live in poverty?**

This varies by country. In the US, 20% have incomes less than half of the average person.

SAV-2. **How much do working people save from their incomes?**

There is great variation in savings, even among people with similar lifetime incomes; many people save none of their income and others save a great deal.

SAV-3**. Do people plan for their old age?**

Only some people plan for their old age. One study showed that one third of people 50 and over had not thought about their retirement needs. Other studies have shown that most workers do not know much about the pensions or benefits they will have in old age; many don't know much about financial matters such as interest rates and inflation.

SAV-4. **Do all employers give their employees a pension when they retire?**

No. One third or more of private companies do not provide their workers pensions.

SAV-5. **Is it easy to predict how much money you will need in old age?**

No. To decide how much they need to save for old age, people need to predict many factors, such as how long they are likely to live, their health, how much interest their savings will gain, and how inflation and interest rates will affect prices.

SAV-6. **Can people be taught to save money?**

People can be taught financial skills and money management. However, studies have shown that classes for workers on money matters have only a small effect on their savings plans and savings behavior.

Order of Q&A Presentation for Team Favoring Social Security Side

SS+1, SS+2, SAV-1

SS+3, SS+4, SAV-2

SS+5, SAV-3, SAV+1

SS+6, SAV-4, SS-1

SAV-5, SAV+2, SS-2

SAV-6, SAV+3, SS-3

SAV+4, SAV+5, SS-4

SAV+6, SS-5, SS-6

Order of Q&A Presentation for Team Favoring Individual Saving Side

SAV+1, SAV+2, SS-1

SAV+3, SAV+4, SS-2

SAV+5, SS-3, SS+1

SAV+6, SS-4, SAV-1

SS-5, SS+2, SAV-2

SS-6, SS+3, SAV-3

SS+4, SS+5, SAV-4

SS+6, SAV-5, SAV-6

Online Appendix B

Study 2 Subgroup Results Breakdown for Percentages of Student Essays Showing each Type of Evidence-based Functional Unit at Least Once in Final Essays

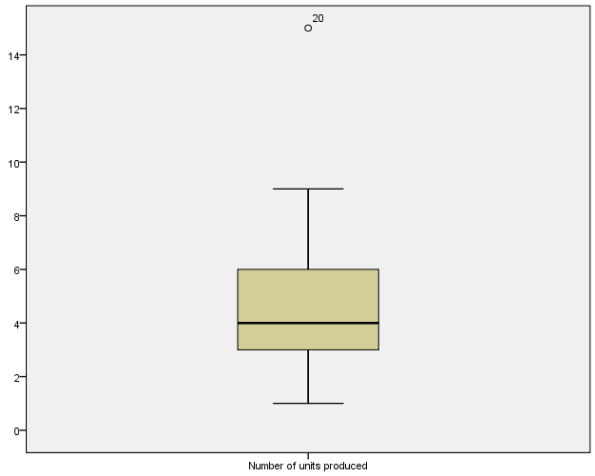
	M+	**O-**	**O+**	**M-**
Incongruent-Evidence-Prompt Group Sequenced presentation (n=11)	**82%**	**73%**	**45%**	**27%**

Incongruent-Evidence-Prompt Group Non-sequenced presentation (n=10)	**80%**	**90%**	**40%**	**30%**
Intervention Group 1 (congruent, prompted) (n=18)	**89%**	**72%**	**22 %**	**06%**
Intervention Group 2 (congruent, non-prompted) (n=19)	**79%**	**68%**	**16%**	**00%**

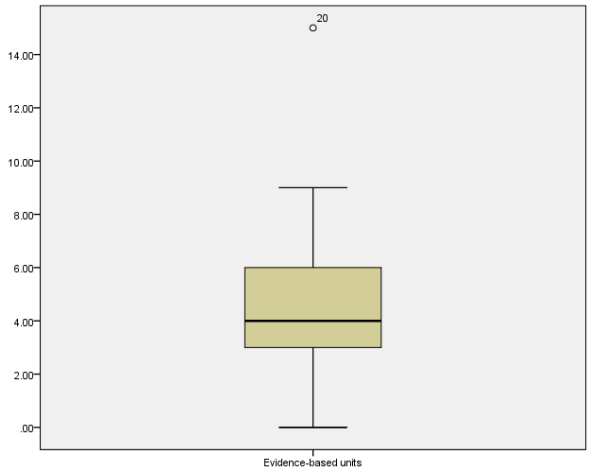
Note. Results are similar for mean values. Data were examined for outliers and these were ruled out. Modal number of such units was zero or one for all groups and subgroups, except M+ and 0- for the Incongruent-Evidence-Prompt Groups, where modes were two.

Online Appendix C

Number of Units at Post-test in Study 1

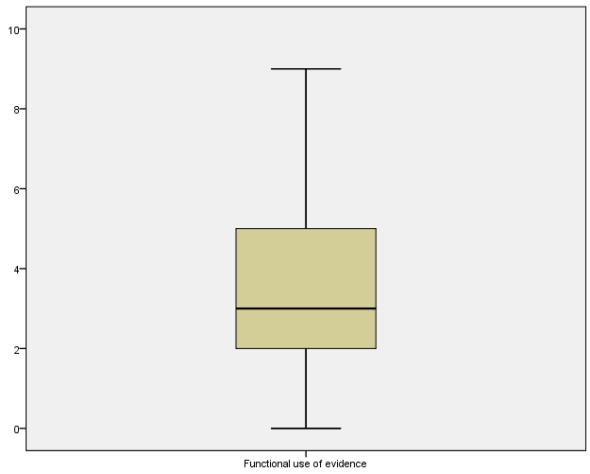


Evidence-Based Units at Post-test in Study 1

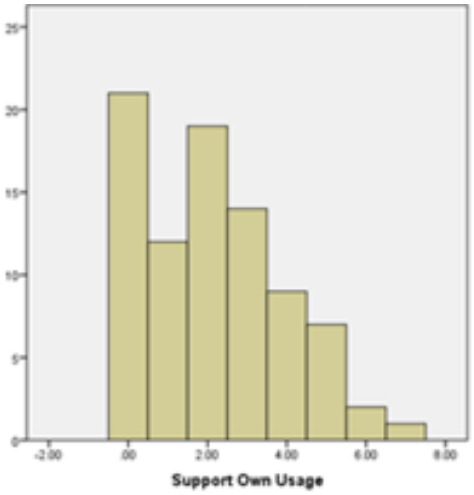


Note. Case 20 (the outlier) is from the control condition, so it worked against Study 1's hypothesis. We have run the analysis without this outlier and the findings between the groups are still the same.

Functional Use of Evidence at Post-test in Study 1



Support-own Usage at Post-test in Study 1



Note. Hoaglin & Iglewicz (1987) demonstrated that the IQR multiplier approach, using 1.5 multiplier that SPSS employs and it is the case here, 50% of the times is inaccurate for identifying outliers. As seen in the histogram above, Case 53 is not an outlier.

Weaken-other Strategy at Post Test in Study 1