[Click for updates](#)

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

Supporting Teachers to Attend to Generalisation in Science Classroom Argumentation

Jonathan T. Shemwell[ab], Kalee R. Gwarjanski[b], Daniel K. Capps[ab], Shirly Avargil[c] & Joanna L. Meyer[b]

[a] College of Education and Human Development, University of Maine, Orono, ME, USA

[b] Center for Research in STEM Education, University of Maine, Orono, ME, USA

[c] School of Education, Faculty of Social Studies, Bar-Ilan University, Ramat-Gan, Israel

Published online: 27 Jan 2015.

International Journal of Science Education, 2015
http://dx.doi.org/10.1080/09500693.2014.1000428

Supporting Teachers to Attend to Generalisation in Science Classroom Argumentation

Jonathan T. Shemwell[a,b*], Kalee R. Gwarjanski[b], Daniel K. Capps[a,b], Shirly Avargil[c] and Joanna L. Meyer[b]
[a]*College of Education and Human Development, University of Maine, Orono, ME, USA;* [b]*Center for Research in STEM Education, University of Maine, Orono, ME, USA;* [c]*School of Education, Faculty of Social Studies, Bar-Ilan University, Ramat-Gan, Israel*

In scientific arguments, claims must have meaning that extends beyond the immediate circumstances of an investigation. That is, claims must be generalised in some way. Therefore, teachers facilitating classroom argumentation must be prepared to support students' efforts to construct or criticise generalised claims. However, widely used argumentation support tools, for instance, the claim-evidence-reasoning (CER) framework, tend not to address generalisation. Accordingly, teachers using these kinds of tools may not be prepared to help their students negotiate issues of generalisation in arguments. We investigated this possibility in a study of professional development activities of 18 middle school teachers using CER. We compared the teachers' approach to generalisation when using a published version of CER to their approach when using an alternate form of CER that increased support for generalisation. In several different sessions, the teachers: (1) responded to survey questions when using CER, (2) critiqued student arguments, (3) used both CER and alternate CER to construct arguments, and (4) discussed the experience of using CER and alternate CER. When using the standard CER, the teachers did not explicitly attend to generalisation in student arguments or in their own arguments. With alternate CER, the teachers generalised their own arguments, and they acknowledged the need for generalisation in student arguments. We concluded that teachers using frameworks for supporting scientific argumentation could benefit from more explicit support for generalisation than CER provides. More broadly, we concluded that generalisation deserves increased attention as a pedagogical challenge within classroom scientific argumentation.

Keywords: *Science teaching; Argumentation; Theory; Evidence; Generalisation; Teacher development; Secondary school*

*Corresponding author. College of Education and Human Development, University of Maine, 330 Shibles Hall, Orono, ME, USA. Email: jonathan.shemwell@maine.edu

> . . . Discovery, whether by a schoolboy going it on his own or by a scientist cultivating the growing edge of his field, is in its essence a matter of rearranging or transforming evidence in such a way that one is enabled to go beyond the evidence so reassembled to additional new insights. (Jerome Bruner, The act of discovery)

Scientific knowledge, as Bruner so eloquently states, is advanced through insights that are extensive in the sense that they have meaning beyond the particular data upon which they are based (Bruner, 1961, p. 90). It follows that claims within scientific arguments, which comprise the vanguard of scientific knowledge, are also laden with extensive meaning. For this reason, scientific argumentation in the classroom naturally involves issues of extensive meaning, including the construction, explication, justification, and limitation of this meaning. For several to-be-stated reasons, we believe that the inclusion of extensive meaning in arguments, for which we adapt the term generalisation, is apt to present a significant challenge for students who are constructing arguments and, by consequence, for teachers who are facilitating this process. If we are correct in this belief, it would be important for instructional materials promoting argumentation to support teachers to negotiate issues of generalisation in claims. However, the bulk of currently available argumentation materials do not provide this kind of support. Therefore, an important question arises as to whether teachers who use these materials would stand to benefit from improved support for dealing with generalisation in classroom argumentation.

We investigated this question in the context of one set of currently available argumentation support materials, a version of McNeill and Krajcik's (2012) claim-evidence-reasoning (CER) argumentation framework, which we argue is representative of available materials in general with respect to our topic. The investigation was a mixed-method study of 18 experienced middle school teachers who were using a published version of CER to help students construct written arguments based on classroom experiments. First, we carefully assessed teachers' baseline understanding and use of generalisation in arguments using CER. Then, we introduced an alternative version of CER, which we designed to better support their thinking about generalisation. We found that when teachers used the curriculum's version of CER, they did not easily recognise the need for student arguments to be generalised. After using the alternative version and reflecting on its use, they appreciated the value of generalising arguments and began to think productively about how to help their students generalise. We concluded that generalisation in argument construction requires more attention than CER and other widely used argumentation tools are currently designed to provide. We further concluded that supporting teachers to help their students construct generalised arguments is but one aspect of a larger imperative that is currently neglected in the literature, namely the need to more explicitly attend to issues of generalisation within classroom argumentation.

The Importance of Generalisation in Classroom Scientific Argumentation

We define generalisation in our context to be the incorporation of meaning in claims extending beyond the particular data to which the claims are related. Within this definition, the term incorporation and, hence, generalisation, can be a verb (i.e. to incorporate or generalise) which would indicate the process of forming, shaping, and justifying extended meaning, or it can be a noun, signifying the presence of extended meaning. We distinguish both noun and verb forms of generalisation from generalisability, which for us is the quality of arguments indicating the degree to which extended meaning of some kind is warranted.

Scientific arguments in the classroom will often demand attention to generalisation because science is fundamentally about theories, and theories are general explanations. Students constructing and critiquing scientific arguments in the classroom will therefore be engaged with generalisation in some way. That scientific arguments are about theory is supported, for instance, by Duschl and Osborne's (2002) proposal that an argument 'addresses the coordination of evidence and theory to advance an explanation, a model, a prediction, or an evaluation' (p. 55). Generality in scientific theory can be seen in Gieres' (1984) description of theories as being definitions of kinds of systems, or Popper's (1959) insistence that theories are universal statements. More precisely, Thagard (1978) located generality in the concept of consilience, which corresponds to the number of different classes of phenomena a theory explains.

One kind of argument for which generalisation is important comprises those which use evidence gathered during an investigation to propose or shape an explanation. This is the kind of argument featured in the present study. Figure 1 provides an example. The figure shows a poster made by a group of eighth-grade physical science students to summarise an investigation. The group rolled marbles of different sizes down a ramp and into a plastic cup, which then slid some distance along a table. The students' procedure and data show that they experimented with different sized marbles; and the larger marbles caused the cup to slide greater distances. Their claim, near the middle of the figure, was 'The cup gets pushed farther from the bigger marbles/bearings.'

According to our definition, the claim in Figure 1 is not much of a generalisation since it is restricted to particular features of the investigation, the sizes of the marbles, and how far the cup was pushed. To generalise more, students would need to extend the meaning of these features to some of the more general ideas they have the potential to represent. For instance, they could form the claim that larger objects push more when they hit something. Going further, they might say that larger objects have more potential to push in them. Pedagogically speaking, these extensions of meaning would represent what Minstrell and Kraus (2005) described as 'opportunities to differentiate between summarising observable results and the conclusions generalised from those results' (p. 482).[1] Similarly, they would begin to fulfil Driver, Newton, and Osborne's (2000) requirement to explain how an observation may be related to other observable events. In this case, the more generalised claims would carry the implication that the marble and cup interactions might apply to other, structurally similar situations.

One pedagogical advantage of either of our proposed generalisations, compared to the original in Figure 1, is that it would better engage the learner with important content ideas potentially embodied by the experiment, such as amount of push or

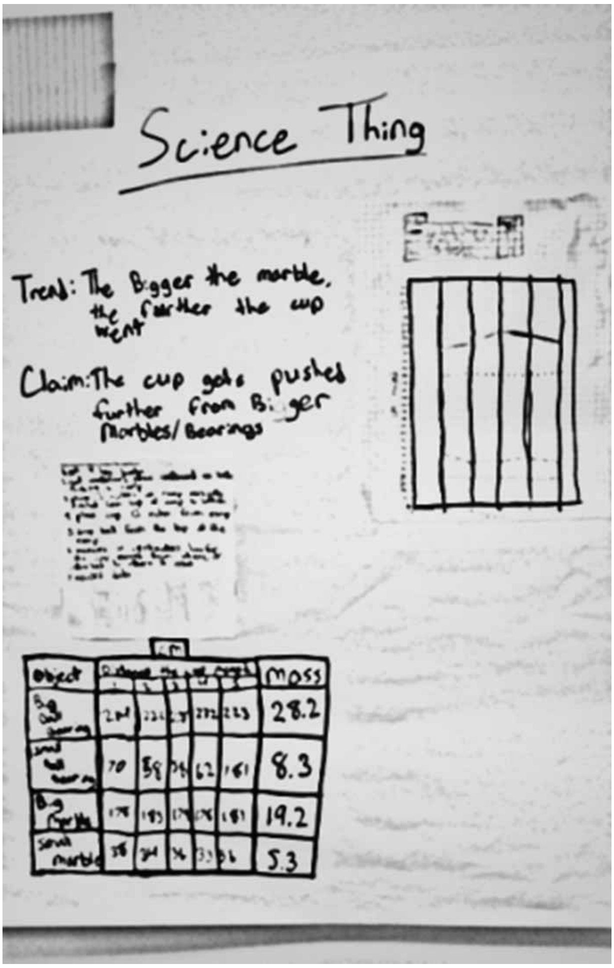

Figure 1. Example of a student argument. The claim is ‘the cup gets pushed further from bigger marbles/bearings’

potential to push. Furthermore, once extended meaning is advanced, there arises a demand to justify it with evidence and reasoning. An example would be the demand to explain why marbles should be taken to represent objects in general. There would also arise the need to limit and otherwise further shape the generalisation. For instance, the author of the claim could ask (or be asked) if it is always true that larger objects push more, and what is meant by larger. In short, many possibilities for learning through and about argumentation would be initiated once some kind of generalisation is put forward. Conversely, if the argument in Figure 1 is not generalised, the claims do not penetrate the surface features of the phenomenon, and there are few opportunities for developing the argument further. Moreover, without generalisation, students could easily come away from the activity with the mistaken idea that scientific claims should be restricted to statements of what was observed and nothing more.

There are several reasons why constructing generalisations like those we propose for Figure 1 would present a significant challenge for students, and, by extension, for teachers who would support them. One is that it can often be difficult for learners to abstract general patterns from particular situations. For example, in studies of analogical transfer, Gick and Holyoak (1983) showed that adults did not easily recognise that an army converging on a city by different roads represented an abstract principle of convergence that could be applied to structurally similar situations. Indeed, many studies have shown that learners are apt to stay at the surface of phenomena rather than think in terms of underlying principles and relationships (Bassok & Holyoak, 1989; Chi, Feltovich, & Glaser, 1981). Complementary studies argue that students are unlikely to induce underlying principles from within particular situations unless they are carefully supported to do so (e.g. Gentner, Loewenstein, & Thompson, 2003; Schwartz, Chase, Oppezzo, & Chin, 2011). A second difficulty is knowing when it is appropriate to advance generalisations within claims, compared to when it is best to keep claims close to the data at hand. For instance, Minstrell and Kraus (2005) described how Minstrell carefully guided students to understand that they should formulate conclusions that were more general than their empirical results, the implication being that this understanding does not come easily and must be carefully supported. Similarly, Maxwell (2013), writing for PhD students in social science, warned that it can be difficult to decide whether research questions should be framed in the particular context of an investigation or in more general terms. A third type of difficulty, pointed out extensively by Walton (2013), is that there are many ways in which students' reasoning can go astray when using evidence to justify or shape generalisations. For instance, students may fail to look for evidence that would limit or falsify a generalisation (Griggs & Cox, 1982; Piburn, 1990). Moreover, students can easily overlook or even reject falsifying information when it presents itself (Chin & Brewer, 1993).

As a final comment on generalisation as a pedagogical issue, we wish to point out that not every scientific argumentation in the classroom will involve students and teachers in concerns of generalisation such as those arising for Figure 1. There can be many scientific arguments for which the level of generalisation is not contested or otherwise in question. As an example, supporting video materials for McNeill and Kracjik (2012) showed a teacher engaging her class in argumentation in which students advanced claims about whether soap and fat were different substances. The argumentation was very productive and meaningful without questioning or shaping the generality of soap and fat as concepts. Similarly, Erduran (2007) pointed out that generalisation is not always an important feature of professional scientists' arguments. She used the example of a biochemist seeking to understand how particular amino acid sequences function within particular proteins, arguing that it would be meaningless for the biochemist to try to establish how the function of the amino acids might generalise. Of course, in both examples, students arguing about soap and a scientist arguing about amino acids, conceptual frameworks would be in place to impart a level of generality to claims. These frameworks would naturally require adjustment with the advancement of knowledge. This idea returns us to the fundamental point motivating the present study, namely that a degree of generalisation is a very important quality within scientific arguments, so this quality will frequently be important when students seek to advance scientific arguments in the classroom.

Generalisation in Frameworks to Support Teaching Scientific Argumentation

Generalisation has been featured to different extents in research frameworks for measuring the quality of students' arguments. Kelly and Takao's (2002) framework provides perhaps the best-known example of including generalisation. Their six-level scheme for analysing written student arguments examined the quality of claims in addition to the degree of support from evidence. The highest quality claim had generalisations that referenced widely accepted definitions and subject matter. Similarly, Furtak and colleagues combined generalisation with support from evidence in rating the quality of students' verbal arguments (Furtak, Hardy, Beinbrech, Shavelson, & Shemwell, 2010). They defined four levels of argument quality. The first three focussed on support from evidence. The fourth level was attained when student claims were explicitly generalised to take the form of a rule. As a contrasting example, the well-known Toulmin Argument Pattern (Erduran, Simon, & Osborne, 2004; Toulmin, 1952) did not address generalisation. The same is true of frameworks developed by Jiminez-Aleixandre and colleagues (Jimenez-Aleixandre, Rodriguez, & Duschl, 2000) as well as Sardá and Sanmartí (2000). These researchers were concerned with other aspects of argumentation than the quality of claims, and hence generalisation. By extension, teaching resources derived from the Toulmin Argument Pattern (e.g. Simon, Erduran, & Osborne, 2006) also do not address generalisation of claims.

Despite its inclusion in some prominent research frameworks such as Kelley and Takao (2002), generalisation is not a conspicuous topic in frameworks and materials explicitly designed to help teachers conduct argumentation in the classroom.[2] Here, we review two widely distributed sets of materials that comprise, to our knowledge, all the currently published resources. These are the CER framework (McNeill & Krajcik, 2012; McNeill, Lizotte, Krajcik, & Marx, 2006) and the science writing heuristic (SWH) (Hand, 2008; Keys, Hand, Prain, & Collins, 1999; Omar, 2008). Both CER and SWH provide elements that highlight the components of a well-constructed argument, including distinct claims. However, the need to consider generalisation of claims is not greatly emphasised in either of them. In CER, an argument is synthesised through explicit attention to each of its three named components, claim, evidence, and reasoning. For our purposes, the most important component is the claim. The primary resource for CER, McNeill and Krajcik (2012), defines a claim as 'a statement that expresses the answer or conclusion to a question or problem' (p. 22). The authors do not address generalisation or any other quality of claims. We speculate that this is done to make the framework as flexible and encompassing as possible. Presumably, it would be up to teachers to manage generalisation and other aspects of claims through their framing and guidance of student investigation and argumentation.

The SWH framework addresses generalisation in arguments to a slightly greater extent than CER. Its materials for supporting teachers briefly state the need for 'finding evidence to support generalisations' (Norton-Meier, Hand, Hockenberry, & Wise, 2008, p. 67). In addition, these materials say that students should 'link arguments to the 'big ideas' of the topic being investigated' (p. 140). However, big ideas are not presented as student-constructed generalisations in relation to data. Rather, they are 'the major concepts the students should leave the classroom with at the completion of the unit' (p. 20). Thus, the materials provide little in the way of guidance or structures for supporting students in negotiating the sorts of generalisations that would embody their own evidence-based understandings of phenomena.

Research on the Teaching of Generalisation in Scientific Argumentation

Few studies have focussed on issues teachers face when supporting generalisation in scientific argumentation, and none have made generalisation the primary focus of their inquiries. Those that have touched on the issue have hinted that generalisation of arguments is not easy for teachers to facilitate. Broadly, the US national report on the 1999 TIMMS Video Study claimed that students in US classrooms had exposure to various forms of evidence, but 'sources of evidence were not frequently linked to larger science ideas to create coherent, connected, in-depth treatment of science content in the lessons' (Roth et al., 2006, p. 21). In a more focussed study, Shemwell and Furtak (2010) illustrated challenges teachers faced when supporting students to construct arguments within classroom discussion. In some cases, teachers tried to restrict students' descriptive language within claims to science vocabulary. We interpret this approach as trying to help students construct claims that were generalised at the appropriate level. Unfortunately, but not surprisingly, it also tended to limit students' opportunities to express and work out their ideas in their own terms. Tabak and Reiser (1999) showed how one middle school teacher worked with students to improve the coherence and generality of their evidence-based claims. At first, students were at an impasse because their claims were isolated statements of fact that were very close to their data. To help the students generate more elaborate claims, the teacher deftly guided students to temporarily set aside considerations of evidentiary support for claims and, instead, work out key relationships between their different empirical statements. The skill with which the teacher facilitated the discourse suggests that providing this kind of support for students is far from easy.

Purpose

To summarise the premise of our study, issues of generalisation are a natural part of scientific argumentation, and negotiating these issues is apt to be a significant teaching and learning challenge. Furthermore, widely available frameworks for teaching argumentation are not very concerned with generalisation. Thus, important questions arise as to how well teachers are supported to work with their students' generalisation of claims when using these frameworks and, potentially, what can be done to improve that support. We addressed these questions in a mixed-method study of teachers as they engaged in professional development activities focussed on scientific argumentation. The teachers were part of a community using a version of the CER framework (McNeill et al., 2006; McNeill & Krajcik, 2012) included in their curriculum (Kolodner, Krajcik, Edelson, Reiser, & Starr, 2009). Our purpose was to find out how well the teachers were supported in working with generalisation of student claims when using CER, and whether they (and others like them) would stand to benefit from improvements to this type of framework.

Method

To accomplish our purpose, we compared the teachers' approach to generalisation when using the curriculum's CER to their approach when using a revised version of CER designed to support generalisation. The comparison did not involve actual teaching using revised or unrevised CER. Rather, it occurred within a professional development context in which teachers constructed arguments or evaluated students' arguments.

Design

The study occurred in three chronological phases designed to develop three related components of quantitative and qualitative evidence. The first was a baseline component in which we examined how teachers thought about generalisation when using the curriculum CER. The second component, intervention, occurred when we observed teachers' use of our revision to the CER designed to support generalisation. The third component was post-intervention, in which we analysed teachers' thinking about generalisation as they discussed their experiences of the intervention.

For the baseline component, we collected data within three different professional development sessions that stretched over a 4-month period. This time-extended data collection was designed to establish the stability of teachers' thinking about generalisation with CER. In the first session, we surveyed teachers' views on generalisation when using CER. In the second, we analysed their critique of student arguments exhibiting different levels of generalisation. In the third, we analysed the level of generality of arguments that teachers themselves constructed when using CER. By contrast, the intervention and post-intervention components were relatively compressed in time, occurring in a single 60-minute session immediately after finishing the last measurement of the baseline component. This compression was necessary to establish a clear linkage between the intervention and changes in teachers' thinking about generalisation.

The intervention component contrasted levels of generalisation achieved by teachers when they constructed arguments in small groups, first using the curriculum's CER, and then using revised CER, with the same data. We predicted, based on CER's low level of support for generalisation, that teachers would generalise very little when using it to construct arguments, and they would generalise more with the revised CER. Meeting both aspects of this prediction would imply that CER stood in need of improvement to better support generalisation.

For the post-intervention component, the entire group of teachers came together to discuss their experiences when using the revised CER framework compared to the original version. We analysed this discussion to gain an understanding of how revised CER influenced teachers' thinking about generalisation. Also, the discussion provided a check on whether differences in generalisation in the intervention component could have resulted from spurious causes, such as the fact that the teachers used the same data twice, instead of the experience of using the revised CER.

Context

The study took place in the context of a 5-year National Science Foundation funded Math and Science Partnership (MSP) involving grades six through nine in rural schools in northeastern USA. The partnership was organised around three grade-level teaching communities, in grades six, eight, and nine. These communities consisted of teachers primarily, together with some university personnel, such as ourselves, including faculty, staff, and graduate students. The study focussed on the eighth-grade community in its second year of curriculum enactment (the third year of the MSP grant).

In the first year of the project, a task force from each community chose a curriculum that its teachers implemented together in subsequent years with the goal of steadily improving teaching and learning. The selection process utilised a modified version of the AAAS Evaluation Tool (AAAS Project 2061, 2002). The eighth-grade community selected *Project-based inquiry science* (Kolodner et al., 2009). It included topics on force and motion, energy, and chemistry and spanned the entire school year. As its name suggests, the curriculum followed a project-based learning format (Krajcik, Blumenfeld, Marx, & Soloway, 1994; Krajcik et al., 1998). Also, it placed a strong emphasis on learning science practices, such as argumentation and explanation, together with science content.

The community was supported in implementing its curriculum and improving teaching in a variety of ways. One form of support came from an online forum for each community in which teachers shared their teaching experiences or suggested options for approaches to teaching. Another form was periodic gatherings in which teachers would work with and reflect on their teaching of the curriculum. These gatherings had several different formats, including a 1-week summer institute each year, and monthly, three-hour community meetings during the school year. The study was conducted during three of these monthly meetings, as part of professional development activities on scientific argumentation and the use of CER.

The curriculum's CER framework. The curriculum supported the construction of arguments from investigations using a version of CER in a worksheet which students used many times in different contexts throughout the year (Figure 2—left). The first section of the sheet (top left of Figure 2) asked students to create a summary of the investigation they performed. Next, students were asked to generate a claim, which the textbook described as being 'a statement of what you understand or conclusion that you have reached from an investigation or set of investigations' (Kolodner et al., 2009, p. DIV58). Clearly, this definition encompassed generalisation in its use of the terms 'conclusion' and 'understanding'. However, the idea that conclusions and understandings have extended meaning was not discussed or implied in any the documentation. This would be consistent with standard CER (McNeill & Krajcik,

Figure 2. (a) The structure of the curriculum's version of CER (Kolodner et al., 2009); (b) The structure of G-CER

2012) and would be understandable given the authors' probable intent to make the framework useful for many different situations, including situations for which generalisation is not an important issue. Following the claim section of the worksheet, students were asked to present evidence that supported the claim. Evidence was described as 'data collected and trends in that data' (Kolodner et al., 2009, p. DIV58). Skipping for the moment the third section of the worksheet, science knowledge, in the fourth and final section (Figure 2—bottom left) students were asked to develop an explanation statement that would bring together their claim, evidence, and science knowledge. The textbook stated that this explanation connected the claim to the evidence and science knowledge in a logical way that would convince someone that the claim was valid.

The third prompt of the curriculum's CER, science knowledge, was in part geared towards helping students to express their reasoning (or warrant) to establish how the evidence supports the claim (i.e. the 'R' in CER). The textbook explained that science knowledge was 'knowledge about how things work [learned through] reading, talking to an expert, discussion, or other experiences' (Kolodner et al., 2009, p. DIV58). It went on to say that 'you use evidence and science knowledge to back up your claim.' However, in contrast to standard CER (McNeill & Krajcik, 2012), the documentation did not include a distinct definition of reasoning as a component of an argument. Thus, there was probably less support for reasoning in the curriculum's CER than in the more current conception of CER.

A second affordance of the science knowledge prompt, more important to us than support for reasoning, was its potential to support generalisation. From this perspective, the prompt for science knowledge would seem to reflect a requirement in US national standards that students should 'make the connection between their results and scientific knowledge' (National Research Council [NRC], 2000, p. 27). In other words, students should understand how results have meaning that extends beyond their own investigation to more generally applicable scientific concepts and principles. However, the supporting material in the textbook did not indicate that connections between empirical results and more general ideas should be made. Thus, while the potential to support generalisation was present in the prompt, this potential was not much developed in the surrounding materials. Nevertheless, our investigation looked closely at teachers' understanding and use of this prompt for generalisation.

Revising CER to support generalisation. Our alternative version of CER (Figure 2—right) was called G-CER (G for general). It included three essential changes from the curriculum's CER. One was to provide an additional section that asked students to name the variables they investigated and what they thought those variables might represent in a more general sense. To use our earlier example of marbles pushing cups (Figure 1), students might write the variable 'size of marble', and then next to it write that it represented something more general, like mass or weight, or how big things are. A second new section called 'trend' asked students to state the results of their investigation in strictly empirical terms (Figure 2—right middle). This section was designed to support students in recognising and describing patterns in their data as a distinct process from thinking about the broader ideas these patterns represented. It was also meant to contrast with the next section, which we re-labelled from 'claim' to 'general claim'. The general claim was intended to be a transformation of the trend into a more generalised statement by utilising the more general forms of the variables in their investigation. The final departure from CER was a reordering of the elements of the prompt. G-CER asked students to summarise their empirical evidence first, then think about their variables, and end with their trend and claim. In the curriculum's version of CER, by contrast, students wrote out their claims before they stated their evidence. This reordering was designed to represent the claim as being a developed idea that would be the endpoint of a generalisation process.

G-CER should not be taken as representing an overall improvement to CER. In fact, a comparison of the two frameworks in Figure 2 will reveal that G-CER does less than the curriculum's CER to support reasoning, a crucial dimension of scientific arguments. G-CER has other potential shortcomings discussed at the end of this paper. Rather, G-CER should be thought of as an alternative version of CER, constructed solely for research purposes, to provide a contrasting level of support for thinking about generalisation. It is the principle of supporting generalisation that G-CER embodies, not the instrument itself, which is meant to be its contribution.

Participants

The participants were the 18 teachers in the eighth-grade community of the MSP. Due to additions and subtractions within the community between the first and second years of implementation, 11 of the teachers were teaching the curriculum for the first time, and 7 were teaching it for the second time. The teachers taught in rural school districts within an area of approximately 70-mile diameter. All taught eighth-grade students, but some also taught students in earlier grades. The teachers ranged in science teaching experience from 1 to 42 years, with a median of 22 years. Half of the teachers had undergraduate degrees in education; the others' degrees were in science, math, or the humanities. All teachers were certified in secondary education. Seven of the teachers were female and 11 were male. All of the teachers were white, non-Hispanic, which reflected the demographic for the rural area in which they taught.

Sources of Data

We collected data at three monthly community meetings in the fall of the school year. There were three types of data for the baseline component of the study, one type for the intervention component, and one type for the post-intervention component. Baseline data consisted of written responses to surveys of individual teachers about classroom argumentation using CER, teachers' written critiques of student arguments, and teachers' written arguments when using CER. Intervention data consisted of written teacher arguments using the revised CER. Post-intervention data consisted of teachers' talk during whole-group discussion as they reflected on using standard CER and revised CER.

Procedures

We collected baseline data at all three of the community meetings. These occurred in September, October, and December. At the December meeting, we also collected intervention and post-intervention data (see Figure 3). Each meeting lasted three hours. Only a portion of each, lasting from 60 to 90 minutes, was dedicated to activities about scientific argumentation, described below. We designed and led these activities as part of our role as university partners within the MSP. The remaining portions were led by other university groups and were dedicated to such things as dinner, administrative time, and diverse professional development topics. Attendance fluctuated; 16 teachers were present for the September meeting, 12 for November, and 17 for December. What follows is a description of professional development activities and data collection procedures for each meeting.

September baseline: surveys. In the first session, we surveyed teachers' views about generalisation in argument. We gave two different surveys, one just before an argumentation activity using standard CER and one just after. At this time, the teachers had been teaching for about 3 weeks, and most had not yet used the CER framework with their students. Teachers responded by computer, taking 10 minutes or less to finish.

We intended the first survey to elicit teachers' most prominent views of the valuable aspects of scientific arguments (see Table 1 for items). We wanted to know whether generalisation was at all salient in teachers' thinking about arguments. Accordingly, we wrote the items to avoid leading teachers to consider generalisation. After finishing the first survey, the teachers took about 30 minutes to conduct an investigation within

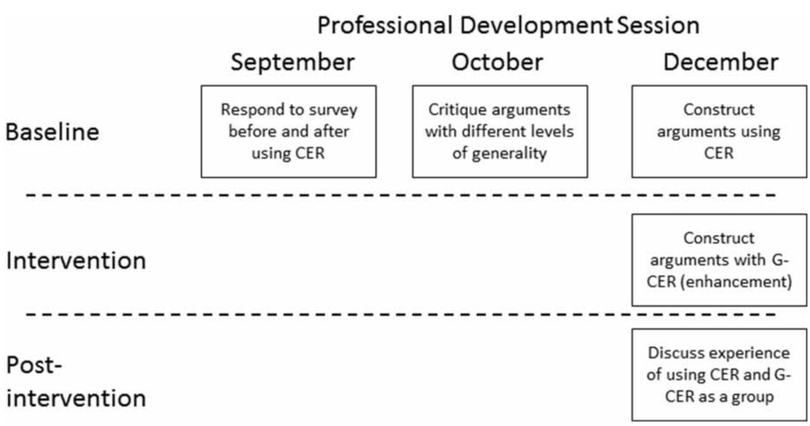

Figure 3. Sequence and timing of data collection for baseline, intervention, and post-intervention

Table 1. Survey items about what teachers valued in science arguments

Item	Description	When asked	
		Before CER	After CER
1	In general, what makes a good scientific argument?	X	
2	Assuming that there are multiple ways that scientific arguments can be good, state as many of these as you can. Which ones are most important for middle school students to know and why?	X	
3	Earlier you wrote about what, in general, makes a good scientific argument. Now that you have done the activity, what would you add to your earlier response?		X
4	How did you use the science knowledge box? How does that support or not support formulation of arguments?		X

their curriculum on sound energy. As part of the investigation, the teachers constructed arguments using CER. Then, they took the second survey (see Table 1). Its purpose was to find out if the experience of using CER had changed teachers' thinking about generalising in arguments. One item asked teachers to explain the purpose of the science knowledge prompt within the curriculum's CER. We included this item to see if teachers thought that science knowledge could be used to help make claims more general.

October baseline: critique of student arguments. At the second session, teachers critiqued student arguments. Most of the teachers had used CER at least once with their students by this time in the school year. We gave each teacher a set of three posters, each of which described an investigation procedure, data, and resulting evidence-based claims. We had obtained the posters from several of the teachers' classrooms from the previous year's use of their curriculum and removed students' names. We carefully selected the posters to feature similar levels of support from evidence, but contrasting levels of generality in claims. Figure 1 shows one of the posters we used. The claims for all three posters are shown in Table 2. We wanted to see if students would notice the variation in generality, and whether they would value the more general claims. We gave each teacher sheets of paper on which they answered the following prompts as they conducted their critique: (1) What are the pluses and minuses of these posters as scientific arguments? (2) Compare and contrast the quality of claims within the arguments. Teachers took approximately 10 minutes to examine the three posters and complete their individual written critiques.

December: final baseline, intervention, and post-intervention. In the third session, teachers began by constructing arguments from data, first using CER (final element of baseline) and then using G-CER (intervention). Then, teachers reflected on the process of using CER and G-CER in discussion (post-intervention). By this time in

Table 2. Variation in generality of claims in student arguments that teachers critiqued

Student argument	Generality	Claim
A	Little	The cup gets pushed farther from bigger marbles/bearings
B	Some	When the angle of the ramp increases, GPE [Gravitational Potential Energy] increases
C	Much	When the steepness increases, gravitational potential energy increases

the year, all of the teachers had used CER at least once with their students; several had used it more than once.

Final baseline: constructing arguments with CER. For the final element of baseline data, teachers worked in small groups to construct an argument together using the curriculum's CER. The data for the arguments had been collected by teachers in the MSP community when they did one of the curriculum investigations in the previous month (i.e. November). In this activity, which was otherwise unrelated to the present study, the teachers compared the distances a toy car travelled when different thicknesses of rubber bands were used to power it. For the ensuing argument-construction activity, we selected one of the resulting data sets and distributed it to each small group of teachers so that all groups had the same data. We then gave groups 20 minutes to develop a scientific argument using the data and the curriculum's CER worksheet (Figure 2—left).

Intervention: constructing an argument with revised CER. Immediately after the teachers finished constructing arguments with CER, we began the intervention component of the study. For this, we gave each group the G-CER worksheet (Figure 2—right) and asked them to repeat the argument construction process, using the same data as before. We left the original CER-based arguments on groups' tables. As before, we gave teachers 20 minutes to construct their arguments. Groups that finished early received a sheet with discussion questions, such as, 'How different were your responses for the two sheets (CER vs. G-CER)?' At the end of the 20-minute period, we collected both CER and G-CER sheets from each group.

Post-intervention: reflecting as a group on using CER and G-CER. After teachers finished constructing arguments with G-CER, we facilitated a whole-group discussion about teachers' experiences using CER and G-CER. We began the discussion by asking two questions: (1) What do you see as the pluses and minuses of G-CER? and (2) Would you be inclined to use something like G-CER, at times, to replace CER? After this initial prompt, facilitation was negligible, as conversation flowed naturally. The discussion lasted 18 minutes and 11 of the 17 teachers contributed in some way. We audiotaped this conversation for transcription and analysis.

Data Analysis

Surveys and critique of student arguments. We consolidated the written survey and argument-critique responses into a table format, organised by teacher. A single researcher began the analysis by open coding all responses for a given question or prompt, to create a list of argumentation-related topics teachers mentioned. Open codes were brief restatements of what the teachers said, similar to what Miles, Huberman, and Saldaña (2014) call descriptive codes. The research team then grouped open codes into categories, which provided second-level codes of interest to the investigation, which we called focus-codes. A single researcher then focus-coded each teacher's response, meaning she examined each response for the absence or presence of each focus code. Finally, she totalled the number of responses giving each focus code for the question or argument-critique prompt.

Teacher-constructed arguments. We analysed small groups' written arguments using CER and G-CER to determine the extent to which their claims generalised beyond the empirical findings of the investigation. A first step in the analysis was to locate and transcribe the most general claims, wherever they appeared on the worksheet. According to the design of both CER and G-CER, these generally occurred in the designated sections for claim or general claim. After we identified the claims, we analysed them both quantitatively and qualitatively. The qualitative analysis focussed on the nature and degree of generalisation within the claims. To facilitate analysis, we assembled all of the claims into a matrix with the different teacher groups in rows, and the type of framework (CER vs. G-CER) in columns. This matrix is presented in the results section, along with analysis of the degree and quality of generalisation in the claims.

A quantitative analysis of teachers' arguments is presented in the results section along with the qualitative data as a way of indexing this information for interpretation by the reader. The analysis used a three-level coding scheme extending from least generalised to most generalised (see Table 3). Notably, the more generalised statements in this table involve teachers' explicit references to scientific principles. These sorts of references were not a requirement of the scheme. We were also open to locally generated constructions we would envision for students (e.g. in place of *energy*, something like *more buildup of tension*). However, not surprisingly, the teachers' generalisations were at a more formal level. For coding procedures, two researchers independently coded all the claims and agreed on all but one of them (96% agreement). They adjudicated the single item to reach consensus.

Teacher whole-group discussion. After transcription, we began the analysis of the whole-group discussion with line-by-line descriptive coding (Miles et al., 2014), in

Table 3. Levels of generality definitions and examples

Level of generality	Definition	Example response
1	No extensive meaning. Remains within the features of the investigation	The thicker rubber band makes the car go farther on average
2	Includes extensive meaning but relates it to specific features of the investigation	A thicker rubber band has more energy and propels the car a greater distance
3	Includes extensive meaning that is independent of the features of the investigation	As you increase the potential energy (elastic or other), you get more kinetic energy released

which a researcher summarised each line of text with a short phrase describing what was said. Next, the researcher used the coded text to develop a chronologically structured descriptive summary of the discussion, which we call a narrative summary. Another researcher then revised the summary, checking it against the original transcript and coding. The finalised summary is provided, together with interpretation and analysis, in the results section.

Results

Results are organised into three sections. The first section provides baseline data from surveys and critique of arguments. The second section compares the final element of baseline data, constructing arguments with CER, to the intervention data, constructing arguments with G-CER. The third section provides qualitative data and analysis from post-intervention, when teachers compared G-CER to CER.

Baseline: Views about generalising using CER

Survey responses about what makes a good scientific argument. Overall, generalisation was not evident in teachers' survey responses about what makes a good scientific argument, either before or after they used CER. Table 4 shows the frequency of different types of response. Before using CER, all of the teachers mentioned evidence as being important. About two-thirds of teachers noted that a claim was necessary, while one-third mentioned reasoning and science knowledge. In an example that combined claim, evidence, and reasoning, a teacher wrote, '[a good argument is] one that makes a clear claim or statement, then gives good evidence that actually supports the claim and uses scientific reasoning to justify.' It was not surprising that claim, evidence, reasoning, and science knowledge were the most common responses, since CER demanded these elements. Additional qualities teachers identified were reliable or repeatable results, rebuttal, leading to more investigations, and collaboration.

The second survey, after teachers practiced with CER, asked teachers how science knowledge contributed to arguments. No teachers mentioned or alluded to

Table 4. What teachers valued in scientific arguments (responses for 16 teachers)

Category	Number of teachers: Before doing CER	After doing CER	Example
Evidence	15	10	A good scientific argument needs to be able to be backed up with proof/data
Claim	9	9	One that makes a clear claim or statement
Science knowledge	5	15	Other science/world knowledge that connects and supports your argument
Reasoning	5	1	Uses scientific reasoning to justify
Conclusion	2	2	Synthesising and wrapping up loose ends, ending
Reliable results	3	0	The proof can be reproduced consistently
Peer review	1	0	Peer review/confirmation
Rebuttal	1	0	Thinking ahead to possible arguments against your own and rebutting them
Leads to more investigations	1	0	Draws a viewpoint that creates new investigations
Collaboration	0	1	Good scientific arguments are done in a collaborative nature drawing upon the experiences of many people or groups of people

generalisation as part of their response. For example, one teacher wrote, 'it [science knowledge] can help support arguments when you have preexisting knowledge that supports your claim, however on its own (without evidence) it can also be manipulated to make false arguments or assumptions seem accurate'. Thus, the teachers did not acknowledge that incorporating science knowledge into arguments could provide theory with which to generalise claims.

While there was no explicit mention of generalisation in arguments either before or after teachers practiced with CER, four responses did state the need for a conclusion, two before using CER and two after. Only one of these elaborated on what a conclusion was meant to provide. This elaboration included the idea that a conclusion involved 'synthesising and wrapping up loose ends' (see Table 4). While synthesis can connote generalisation, its use in this case seemed to imply summarisation. This was as close as teachers came to addressing the issue of generalisation.

Teachers' critique of student arguments. In contrast to survey responses, when teachers critiqued student arguments, at least some of their thinking seemed to be about the value of generalisation. This occurred when five of the teachers mentioned the importance of using appropriate scientific vocabulary. In two of the five cases, teachers talked about vocabulary as if it were a means of expressing generalisation. Table 5 shows one case, when a teacher talked about the need to use vocabulary such as potential energy in the claim instead of simply restating the empirical result. The remaining three teachers simply stated that appropriate vocabulary or terms were needed; so it is unclear whether they may have been implicitly thinking about generalisation when making these comments.

Evidentiary support for claims was the most frequently mentioned category when critiquing arguments (see Table 5). Reasoning was less frequently mentioned in critique than in survey responses. This could have been because reasoning was not particularly explicit in the example posters. A new issue arising in critique, compared to survey responses, was the need for clarity, including the need for a clear statement defining the claim and carefully defined units. There were no indications that either of these critiques was related to the value of generalisation.

Baseline to Intervention: Constructing arguments using CER and then G-CER

Table 6 shows the extent to which teachers generalised when they constructed arguments, first using the curriculum's CER and then using G-CER for the same data. The left hand column, when teachers used the curriculum's CER, represents the third and final element of the baseline data. It shows that, except for one group (B), teachers scarcely generalised when using CER. Instead, most groups expressed their results in strictly empirical terms. Group C was typical when they said, 'If you want to go further, use a thicker rubber band.' All of the other groups made structurally identical statements, except group B, which talked about the thicker rubber band having more energy. When using the G-CER on the second pass with the data, most groups generalised their claims much more. For instance, Group C claimed that with greater rubber band thickness there would be more tension (presumably when stretching or winding) and therefore more force upon release. One of the groups, group E, did not finish in time to produce a claim using G-CER. A member of this group wanted to talk about difficulties in his classroom unrelated

Table 5. What teachers noticed when critiquing student arguments (responses for 11 teachers)

Criteria	Number of Teachers	Example
Clarity	6	[Poster B] is clear about the claims
Evidentiary support	8	All claims were supported with the data presented, either through the table or graph
Reasoning	1	[Poster B] has a data table but no clear claim or reasoning. [poster C] has some solid (relatively speaking) scientific reasoning
Scientific vocabulary	5	The claim that [poster A makes] is not very specific. The language of the claim does not use any of the key vocabulary such as potential energy. It is essentially just a restatement of the trend they wrote above
Standardised units	2	[One group] uses measurement of the ramp in inches while the other group uses books, which is not a measurable unit

Table 6. Generalisation in claims produced by teachers using CER and G-CER (17 teachers in 5 groups)

Group	CER		G-CER	
	Generality level	Claim	Generality level	Claim
A	1	The rubber band size changes the distance the car travels; the thicker rubber band will travel farther	3	As you increase the potential energy (elastic or other), you get more kinetic energy released
B	2	A thicker rubber band has more energy and propels the car a greater distance	2	A thicker rubber band has more elastic potential energy, which transfers to more KE which results in greater distance
C	1	If you want the car to go farther, use the thicker rubber band	2	Increasing the tension as you wind something with elasticity, the thicker the material is will correlate to increased force upon release
D	1	The size of the rubber band changes the distance the car will travel. The thick rubber band travels further (with the same number of spins) than the thin	3	The more energy can be stored (potential) the more available kinetic energy when released. Kinetic energy increases with the increase in potential energy
E	1	The thicker rubber band makes the car go farther on average	–	–

to argumentation, so the group mostly discussed these issues instead of completing G-CER.

Teachers' claims when using the G-CER (see Table 6) clearly reveal how readily they understood that rubber band thickness could be related to larger science ideas in their curriculum such as winding tension, force of release, and energy. Why did they not make these generalisations when using CER? One explanation, consistent with results of the surveys and critique of student arguments, is that the teachers using CER did not realise that that it would be appropriate to generalise. This explanation accords with the fact that nothing in CER directly suggested that generalisation was beneficial or desirable. Of course, CER did have the prompts to describe relevant science knowledge and then use this knowledge to create an explanation. However, the teachers did not seem to interpret these prompts as invitations to generalise. Another explanation, complementary to the first, is that the teachers did not have a good idea of how to generalise when using CER. By this line of reasoning, the suggestion in CER to use science knowledge was too vague; it did not suggest productive avenues for generalisation. By contrast, G-CER enabled teachers to generalise because it focussed their thinking on how particular variables, such as rubber band thickness, might connect to larger ideas such as elastic tension or energy. However, the data at hand, particularly the ease with which the teachers generalised with G-CER, provide perhaps more support for teachers not thinking that they should generalise than not knowing how. Had teachers not known how to generalise, the increase in generalisation would probably have been less dramatic. More importantly, either possibility, not knowing to generalise or not knowing how, indicates a need to improve support for generalisation in CER.

A confounding alternative to both explanations immediately above is that having a second opportunity to construct arguments made it easier to generalise, perhaps because teachers were more familiar with the data the second time around. Analysis of the post-intervention data, presented next, checked for this and other spurious causes for the increased generalisation with G-CER.

Post-intervention: Whole-group discussion of generalisation using CER and G-CER

Post-intervention evidence and findings consist of a qualitative analysis of the teachers' whole-group discussion as they reflected on the differences between using CER and G-CER. In what follows, we present our narrative summary of this discussion, divided into three temporal segments corresponding to three broad topics discussed by teachers. In segment 1, several teachers voiced appreciation for how G-CER explicitly supported generalisation compared to CER. In segment 2, some teachers expressed reservations about generalising with G-CER and they began to generate their own ideas for supporting students to generalise claims. In segment 3, one teacher wrapped up the discussion by summarising how G-CER helped her group generalise more than CER did.

We present each segment's narrative summary, then we provide a brief findings section containing analysis and claims about the meaning of the discussion in that segment. Then, after the third segment, we provide an overall summary of findings for post-intervention.

Narrative summary of segment 1: appreciation for G-CER. After two participants made some general remarks, one teacher, Mary, began the conversation by comparing the affordances for generalisation in G-CER and CER. She explained that with G-CER, 'after stating your independent variable you had to answer what do you think it represents. We thought that led them more to think about what they are doing.' Another teacher, Megan, briefly responded to Mary that generalisation might be difficult, saying 'it might be hard to get that.' Next, Karl shared that students tended not to see that their explanations at the bottom of the CER worksheet should extend beyond the summary of results. Rather, students tended to restate their results.

Eddie immediately agreed with Karl. He went on to say that students' CER explanations should come from rewriting and polishing their initial results summaries, which is hard for students to understand. Two other teachers, Gayle and Wilson, responded to Eddie by pointing out that the explanation section at the bottom of CER should be used for generalisation. Gayle also pointed out that CER did not explicitly prompt for generalisation: 'I thought the final box on the [CER] was trying to make a more general, make your claim in a more general sense, but it doesn't say that. The new [G-CER] actually asked for it.' Gayle went on to say that she valued G-CER's explicit support for generalisation:

> I think having the 'what is your independent variable' and 'what does it represent' helps them generalise, which will help them down the road. That was the most valuable part. You are not saying a claim specific to this situation; they are talking about KE and PE. I was very happy about that.

Following this statement, another teacher, Wilson, reinforced Gayle's emphasis on the importance of generalisation, saying 'This experiment is about the concept you are trying to learn about.' Wilson also reiterated Gayle's point that CER science knowledge prompt could be used to support generalised claims.

Findings for segment 1: appreciation for G-CER. The discussion in segment 1 shows that several teachers recognised the value of generalisation using G-CER. For instance, Wilson's parsimonious 'This experiment is about the concept' was a strong statement of why generalisation is important in argumentation. Two of the teachers, Gayle and Mary, also explicitly praised the way that the G-CER structure supported generalisation. This praise suggests that they saw G-CER's support as the reason they generalised more during the intervention phase. Similarly, Karl's observation that students using CER tended to restate their empirical results drew attention to the increased support for generalisation in G-CER.

Interestingly, for the first time in the study, some of the teachers saw how CER could be amenable to generalisation, as when Gayle pointed out that the explanation section of CER should be used for generalised claims. Not all of the teachers who spoke recognised this affordance, as Eddie's focus on rewriting and polishing suggests that he was not, at that time, connecting CER explanation to generalisation. His view was consistent with the fact that most small groups had not generalised earlier in the evening when using CER (i.e. in the final baseline measurement). While Gayle's (also Wilson's) rejoinder to Eddie clearly showed that she realised that explanation in CER could include generalisation, she also reflected that CER did not explicitly ask students to generalise. We think she probably saw this omission as a potential deficiency, as she went on to talk about the value of the more explicit supports within G-CER.

Finally, Megan's brief interjection that generalisation with G-CER 'might be hard to get' suggests she was thinking about how difficult it might be for students to construct generalisations on their own. She reinforced this idea at the beginning of the next segment of the conversation.

Narrative summary for segment 2: reservations about G-CER. Picking up where the first segment left off, Megan countered Gayle's praise for the variable/what it represents format (quoted in previous segment), saying that having students write what variables stand for 'won't always tell me what they know'. She then elaborated, saying that students would need practice to go from something like rubber band thickness to a more general idea of what it represented. Henry, who had not yet spoken, agreed, referring to a 'jump' between rubber band thickness and potential energy. Then, Henry began to talk about revisions to draft claims, and he introduced the idea of helping students walk backward from more general ideas to specific situations. Eddie extended Henry's idea by constructing a possible prompt or question to ask students: 'Some type of statement, why did you choose this variable? Why do you think it would have the effect?' Then Karl picked up the theme. He explained that G-CER did not ask students to talk about what they knew, or what they were learning in the unit:

> That might be something that is missing; it doesn't say what do you already know? It doesn't ask, 'What do you already know about KE and PE?' It doesn't say, 'What have we been reading and talking about in this section?' To say maybe 'I should be thinking about PE, not about elastic thickness and spins.'

Henry agreed with Karl, and suggested that it would help to ask students to think about a scientific concept from the lesson or unit. Eddie also agreed, stating that G-CER was missing the science knowledge prompt featured in CER. (The conversation moved next to a discussion of the amount of writing required in CER and G-CER. Talk continued on this topic for about four minutes before turning back to wrap up about generalisation in G-CER, presented as phase 3.)

Findings for segment 2: reservations G-CER. The discussion in segment 2 shows the teachers more deeply engaged in thinking about how students need to be supported in order to generalise successfully. This engagement began with Megan's concern about what students would understand when using the G-CER format to generalise. Her concern, together with Henry's reservation about the 'jump', recognised the danger of hollow generalisations in which students would simply label what they observed without making meaningful connections. It is unclear from Megan's remarks whether she thought it was unrealistic for students to generalise as a rule, or whether the way G-CER supported generalisation was unrealistic. By contrast, Henry, Karl, and Eddie, by thinking of ways of supporting students to generalise, signalled that they thought generalisation was a valuable and potentially attainable goal. Their thinking about supporting generalisation extended beyond the G-CER structure. For instance, Henry suggested working backward from generalised claims to the specific data in hand. The novelty of this idea and the clear attention to student thinking by Megan, Karl, Eddy, and Henry are evidence that the experiences of using G-CER and reflecting on its use in the whole-group discussion had an expanding influence on these teachers' thinking about student generalisation. Also, Eddie evidenced a clear advancement in his understanding of the potential of CER when he acknowledged that its science knowledge prompt might be an affordance for generalisation. In this, he extended Gayle and Wilson's insights from segment 1 about the fact that explanation should include generalisation.

Narrative summary for segment 3: final comparison of CER and G-CER. The facilitator said that the session needed to wrap up and asked for any last thoughts. After a pause, Gabby, who had not yet contributed to the conversation, referred to the generalisation structure of G-CER and why she thought it was helpful. She reflected on her own thinking when completing the G-CER sheet, and the fact that her group made connections to more general ideas that they had not made when using CER:

> Where this one [CER] is so vague because they are using it [lists sections; reads evidence] and the kids are like okay, and they copy their data chart. When we were doing this [G-CER] we went right through very quickly and oh wow look at those connections. This is a much more user-friendly way of getting at the concept versus a thick rubber band makes the car go farther. Which I asked [university physics professor] do kids really need to know this?

The facilitator then thanked the teachers, bringing the discussion to a close.

Findings for segment 3: final comparison of CER and G-CER. Gabby's summary statement explained how CER did not convey support for generalisation, while G-CER enabled meaningful connections to science ideas. Her statement was an explicit description of the causal effect of the G-CER structure on the teachers' approach to constructing arguments. We concluded from this statement, together with similar but less explicit statements in segment 1 by Mary, Gayle, and Wilson, that the higher levels of generalisation achieved in the intervention part of the study (see Table 6) did, in fact, result from using G-CER and not from some spurious cause, such as constructing the same argument twice.

Referring to students' use of CER, Gabby said that they would 'copy their data chart', meaning that students tended not to generalise beyond the immediate circumstances of their investigations. This remark and her overall summary showed that she appreciated the value of students making generalisations in arguments. In this sentiment, she had been joined at one time or another by all but one of the teachers who took an active part in the discussion. Only Megan did not say anything in favour of generalisation. Thus, we interpreted Gabby's statement to reflect an overall trend in the discussion that teachers recognised and appreciated the role generalisation in arguments. By contrast, when using CER, teachers had not recognised that empirical results could and should be argued in a more general sense.

Summary of findings for post-intervention. The primary purpose of the analysis of the whole-group discussion—and this was irrespective of what caused teachers to generalise when using G-CER—was to show in detail how teachers' thinking could benefit from the opportunity to engage in and reflect on generalisation in arguments. Most fundamentally, this opportunity enabled the teachers to recognise that it was appropriate for students to try to connect their empirical investigations with the cars' rubber bands to more general ideas. Furthermore, explicit prompts and structural support within G-CER provided teachers with a model for thinking about how to support students in the generalisation process. Not all teachers were comfortable with the G-CER model, but this too was productive, because teachers' criticisms provided a jumping off point for generating their own ideas for supporting generalisation. Finally, for Gabby, and we suspect for others, the experience of generalising sparked new insights into the purpose of the empirical investigations embedded in the group's project-based curriculum. She realised there can and should be user-friendly ways of moving from student's hands-on experiences to the broader ideas those experiences represent. Taken together, these findings provide the real value and contribution of the present study, which is to illustrate how teachers can benefit from increased support for explicitly considering generalisation in students' scientific arguments, compared to what is currently available within argumentation tools like CER.

Discussion

Our overall finding was that using CER on its own did little to support teachers in thinking about generalisation in arguments, compared to when teachers used G-CER. This finding is backed by four different results. First, generalisation was not a quality of good arguments that any of the teachers mentioned in their survey responses, either before or after using CER. Second, the teachers did not recognise or value generalisation as a criterion when critiquing student arguments. Here we acknowledge that some teachers mentioned the need to use appropriate vocabulary, suggesting that they intuitively valued generalisation. However, only two of the teachers' ideas about vocabulary could be associated with generalisation. Third, the teachers did not generalise when constructing their own arguments using CER, despite the fact that their data could support generalisation. Fourth, when teachers used G-CER, they generalised much more in their own arguments, and when they reflected on using G-CER, they began to value student generalisation and appreciate why it was important.

How applicable are these findings to teachers and situations outside of our study? We think the combination of sample and findings provide enough grounds to seriously question whether teachers using CER, broadly speaking, are sufficiently supported to attend to students' generalisation of arguments. This question would extend to teachers using non-CER frameworks that provide little explicit support for generalisation. A limitation of applicability would stem from the fact that, although the teachers were reasonably familiar with CER, they were not experts. Perhaps teachers who have used CER for many years would show a different approach to generalisation of arguments. Nevertheless, trends in reform foretell a dramatic increase in the number of teachers worldwide who will engage students in scientific argumentation (NGSS Lead States, 2013; NRC, 2012; OECD, 2010; Osborne & Dillon, 2008); so presumably, there will be a growing number of teachers whose experience levels with argumentation frameworks resemble those of teachers in the present study.

Participants in the study were middle school teachers and therefore science teaching generalists. Perhaps our results would have been different for teachers of higher grades or at university, given those teachers' stronger discipline and content affinities. It could be argued, for instance, that disciplinary specialists could better envision the potential of specific situations to generalise to theories within their disciplines. On the other hand, we do not think that specialists would be more aware than generalists that students need to be supported in generalisation. Similarly, specialists would not be expected to have superior knowledge or skills for supporting classroom argumentation, broadly speaking. Consequently, we would expect support for generalisation in argumentation to benefit content specialists, even if it may be less urgently needed for them than for generalists like the teachers in the present study.

The order of the claim, evidence, and reasoning prompts in the version of CER presented here suggested that students should formulate their claims before stating their evidence. G-CER reversed this order, asking students to summarise their evidence and then formulate claims. We made this change because we thought it impractical for students to make claims before thinking about what the data were telling them. By adjusting the order, we asked students to generalise more inductively, from specifics to generalities. We think that this method may better support teachers and students to develop arguments containing more generalised theory. Developing ideas through inductive generalisation does not exactly align with the philosopher's ideal of science, which is hypothetico-deductive (Popper, 1959), placing claims prior to data. However, real science is often inductive (Chinn & Mathothra, 2002; NRC, 2012). Moreover, processing information inductively, by extracting latent structure from among multiple similar situations, is very effective for constructing meaningful generalisations. This effectiveness has been shown in the general psychological literature (Gick & Holyoak, 1983), as well as learning research in mathematics (Rittle-Johnson & Star, 2007), business (Gentner et al., 2003), and science (Schwartz et al., 2011; Shemwell, Chase, & Schwartz, 2015). Here it bears emphasising that we are not proposing that every argument's claim should be arrived at through induction. Rather, our point is that the question of whether claims are prior to evidence, or vice versa, is an important one to consider when designing or selecting argumentation support tools.

Some teachers in the study expressed concern that students might construct hollow generalisations if they were pushed to generalise too far and too fast. We were similarly concerned. We worried that we had unwittingly misrepresented generalisation as making swift and sure connections from empirical results to high-level science ideas. One of the foundations for our worry was the large jump in generalisation that some of the small groups made when using G-CER. For instance, several of the groups went from rubber band thickness to elastic potential energy. We had envisioned them going from rubber band thickness to something more proximate to students' experiences and forms of expression, such as the effort it takes to stretch something. The fact that we did not achieve this vision underscores the caveat voiced when first introducing G-CER that this instrument was not meant to represent an improvement to CER or otherwise act as an exemplar. Rather, G-CER was a rough and ready research tool for showing that it is both possible and beneficial to better support teachers in thinking about generalisation than was done in CER. Therefore, we do not recommend that G-CER be directly applied to science classrooms. At best, G-CER should be viewed as an early prototype from which future tools can be built to support generalisation in arguments. The real contribution of G-CER is the idea motivating its invention: that generalisation should receive explicit consideration and support in teaching of and through scientific argumentation.

Notwithstanding our caution about G-CER, a natural place to improve upon generalisation in classroom scientific argumentation would be to augment frameworks already in use in the classroom (e.g. McNeill & Krajcik, 2012; Norton-Meier et al., 2008). Our approach, driven by the exigencies of research, was to modify the existing framework structurally, by providing separate sections for empirical and general claims, and adding prompts for participants to state the more general ideas the variables in their experiments represented. The modifications came at the cost of removing support for other important aspects of argumentation, for instance, support for reasoning. Perhaps a version of G-CER could be created that would improve support for elements like reasoning and also integrate the field's growing knowledge of how to support generalisation (e.g. Schwartz et al., 2011, Shemwell et al., 2015). While awaiting such changes, a useful step could be to begin focusing on generalisation in argumentation without structural changes to frameworks and instruments. This could be done by changing supporting materials for existing frameworks and revising associated professional development. For instance, in the present study's version of CER, supporting documentation could describe how the curriculum's scientific knowledge prompt could be used to support generalisation, along with reasoning. The simplicity of such a modification underscores the essential point of the present study. What are needed most are not necessarily new or better tools or procedures, but rather attention to the question of how teachers and students can think about and approach generalisation within arguments.

Conclusion

We began by pointing out that generalisation, or extended meaning, within claims is naturally an important aspect of classroom argumentation, and that support for engaging students with issues of generalisation appears to be lacking in prominent argumentation teaching tools. We then provided evidence showing how a group of teachers using one tool, a version of CER, attended very little to generalisation until they experienced the alternative version we provided. We argued from this evidence that teachers at large would probably stand to benefit from better support for thinking about generalisation in students' arguments than is currently available, at least on a wide scale. Taken together, these findings and arguments point to a larger necessity for science education, which is to acknowledge and attend to questions about generalisation of claims within classroom argumentation, including questions of how to generalise, how much, why, and with what justification.

Finally, it is worth noting that generalisation is but one aspect of the broader issue of evaluating or building theory in classroom argumentation. Since arguments are fundamentally explanations of theory in relation to evidence, questions about the nature of theory incorporated into claims are fundamental for any student constructing an argument. This is not to say that every student argument should be filled to its brim with theory, or that every argumentation process should focus on theory. It is to say, rather, that the quality of theory within claims should be an explicit consideration in classroom argumentation, on par with considerations of evidence and reasoning.

Acknowledgement

This work was supported by a grant from the National Science Foundation (DRL-0962805). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the granting agency.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Notes

1. Following the lead of Peirce (1878), it is worth pointing out that the construction of extended meaning is not necessarily an outward flow from particular data to generalised claims. The process could instead be thought of as invoking existing general patterns to explain a particular set of observations. Our example generalisation, 'larger objects can push more when they hit something' implies that the particular case (a cup) could be extended to represent a more general concept, (something, or any massive thing). However, this example could be thought of as mapping a concept already in memory (something) to the particular case (a cup).
2. For an example of a framework that is not expressly designed to support classroom argumentation that places significant emphasis on generalisation, see Gowan's Vee (Novak & Gowan, 1984).

References

AAAS Project 2061. (2002). *Middle grades science textbooks: A benchmarks-based evaluation*. American Association for the Advancement of Science. Retrieved from http://www.project2061.org/publications/textbook/mgsci/report/mgbooks.htm

Bassok, M., & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 153–166.

Bruner, J. (1961). The act of discovery. *Harvard Educational Review*, *31*, 21–32.

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152.

Chin, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, *63*(1), 1–49.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*(2), 175–218.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, *84*(3), 287–312.

Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, *38*(1), 37–41.

Erduran, S. (2007). Methodological foundations in the study of argumentation in science classrooms. In S. Erduran & M. P. Jimenez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 47–69). Dordrecht: Springer.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education*, *88*(6), 915–933.

Furtak, E. M., Hardy, I., Beinbrech, C., Shavelson, R. J., & Shemwell, J. T. (2010). A framework for analyzing evidence-based reasoning in science classroom discourse. *Educational Assessment*, *15*(3–4), 175–196.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, *95*(2), 393–405.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–38.

Giere, R. N. (1984). *Understanding scientific reasoning* (2nd ed.). New York, NY: CBS College.

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, *73*(3), 407–420.

Hand, B. (2008). *Science inquiry, argument and language: A case for the science writing heuristic*. Rotterdam: Sense.

Jimenez-Aleixandre, M. P., Rodriguez, A. B., & Duschl, R. A. (2000). 'Doing the lesson' or 'doing science': Argument in high school genetics. *Scence Education*, *84*(6), 757–792.

Kelly, G. J., & Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science Education*, *86*(3), 314–342.

Keys, C. W., Hand, B., Prain, V., & Collins, S. (1999). Using the science writing heuristic as a tool for learning from laboratory investigations in secondary science. *Journal of Research in Science Teaching*, *36*(10), 1065–1084.

Kolodner, J. L., Krajcik, J. S., Edelson, D. C., Reiser, B. J., & Starr, M. L. (2009). *Project based inquiry science (Diving into science)*. Armonk, NY: It's About Time.

Krajcik, J. S., Blumenfeld, P. C., Marx, R. W., & Soloway, E. (1994). A collaborative model for helping middle grade science teachers learn project-based instruction. *The Elementary School Journal*, *94*(5), 483–497.

Krajcik, J., Blumenfeld, P. C., Marx, R. W., Bass, K. M., Fredricks, J., & Soloway, E. (1998). Inquiry in project-based science classrooms: Initial attempts by middle school students. *Journal of the Learning Sciences*, *7*(3–4), 313–350.

Maxwell, J. A. (2013). *Qualitative research design: An interactive approach*. Thousand Oaks, CA: Sage.

McNeill, K. L., & Krajcik, J. S. (2012). *Supporting grade 5–8 students in constructing explanations in science: The claim, evidence, and reasoning framework for talk and writing*. Boston: Pearson Education.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences*, *15*(2), 153–191.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook*. Thousand Oaks, CA: Sage.

Minstrell, J., & Kraus, P. (2005). Guided inquiry in the science classroom. In M. S. Donovan & J. D. Bransford (Eds.), *How students learn: Science in the classroom* (pp. 475–513). Washington, DC: The National Academies Press.

National Research Council. (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. Washington, DC: The National Academies Press.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: National Academies Press.

Norton-Meier, L., Hand, B., Hockenberry, L., & Wise, K. (2008). *Questions, claims, and evidence*. Arlington: National Science Teachers Association.
Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. London: Cambridge University Press.
OECD. (2010). *PISA 2009 results: What students know and can do—student performance in reading, mathematics and science* (Vol. I). Pisa: Author.
Omar, S. (2008). *The science writing heuristic as a tool: Professional development of in-service teachers*. Saarbrücken: Verlag.
Osborne, J., & Dillon, J. (2008). *Science education in Europe: Critical reflections* (A Report to the Nuffield Foundation). London: Nuffield Foundation.
Peirce, C. S. (1878). Deduction, induction, and hypothesis. *Popular Science Monthly*, *13*, 470–482.
Piburn, M. D. (1990). Reasoning about logical propositions and success in science. *Journal of Research in Science Teaching*, *27*(9), 887–900.
Popper, K. (1959). *The logic of scientific discovery*. New York: Routledge Classics.
Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, *99*(3), 561–574.
Roth, K. J., Druker, S. L., Garnier, H. E., Lemmens, M., Chen, C., Kawanaka, T., ... Gallimore, R. (2006). *Highlights from the TIMSS 1999 video study of eighth-grade science teaching*. Washington, DC: National Center for Education Statistics.
Sardá, J. A., & Sanmartí, P. N. (2000). Enseñar a argumentar científicamente: un reto de las clases de ciencias. *Enseñanza de las Ciencias*, *18*(3), 405–422.
Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, *103*(4), 759–775.
Shemwell, J. T., Chase, C. C., & Schwartz, D. L. (2015). Seeking the general explanation: A test of inductive activities for learning and transfer. *Journal of Research in Science Teaching*, *52*(1), 58–83.
Shemwell, J. T., & Furtak, E. M. (2010). Science classroom discussion as scientific argumentation: A study of conceptually rich (and poor) student talk. *Educational Assessment*, *15*(3–4), 222–250.
Simon, S., Erduran, S., & Osborne, J. (2006). Learning to teach argumentation: Research and development in the science classroom. *International Journal of Science Education*, *28*(2–3), 235–260.
Tabak, I., & Reiser, B. J. (1999). *Steering the course of dialogue in inquiry-based science classrooms*. Paper presented at the annual meeting of the American Educational Research Association, Montreal.
Thagard, P. R. (1978). The best explanation: Criteria for theory choice. *The Journal of Philosophy*, *75*(2), 76–92.
Toulmin, S. E. (1952). *The uses of argument*. Cambridge: Cambridge University Press.
Walton, D. (2013). *Argumentation schemes for presumptive reasoning*. New York: Routledge.