Check for updates

Posing New Researchable Questions as a Dynamic Process in Educational Research

Gabriel J. Stylianides[1] · Andreas J. Stylianides[2]

Abstract

In this paper, we argue that posing new researchable questions in educational research is a dynamic process that reflects the field's growing understanding of the web of potentially influential factors surrounding the examination of a particular phenomenon of interest. We illustrate this thesis by drawing on a strand of mathematics education research related to students' justification schemes that has evolved rapidly during the past few decades. Also, we reflect on the possible boundaries of the domain of application of the thesis, and we hypothesize that it would apply equally to other strands of educational research. To support this hypothesis, we briefly consider how the thesis would be applicable in two additional research strands. We conclude by elaborating on three important implications of the thesis: (1) as new potentially influential factors about the phenomenon of interest are identified, findings from past research that had not accounted for those factors might prove to be insufficient or be put into question; (2) there are increased methodological challenges for researchers as they seek to design new studies that pay due regard to research advances about all relevant and potentially influential factors surrounding the phenomenon of interest; and (3) as a wider range of potentially influential factors get discovered and considered about a particular phenomenon, research knowledge becomes not only more refined, and presumably more accurate, but possibly more fragmented too.

Keywords Educational research · Justification schemes · Proof · Researchable questions

---

The two authors contributed equally to the preparation of this paper. Part of an earlier version of the paper will be presented at the 14th International Congress on Mathematical Education (Shanghai, China).

✉ Gabriel J. Stylianides
gabriel.stylianides@education.ox.ac.uk

Andreas J. Stylianides
as899@cam.ac.uk

Extended author information available on the last page of the article

🌱 Springer

Introduction

In every empirical study in educational research, the researcher needs to make decisions on how to deal with factors that can possibly influence the phenomenon under examination. In particular, when posing a new research question, the researcher needs to make decisions about whether (a) to ignore certain factors that are deemed unimportant, (b) to document other factors heavily or using a "light touch" approach depending on their expected or presumed influence on the phenomenon under examination, or (c) to modify the research design so as to make non-applicable certain potentially influential factors that are nevertheless of no special interest to the new research. In posing a *researchable question* (i.e. a question that can be investigated empirically), decisions about how to deal with the whole web of factors that can possibly influence the phenomenon under examination need to be made carefully, transparently, and in a justifiable way, drawing on relevant theoretical or conceptual frameworks and associated empirical findings. If proper handling of potentially influential factors is not possible, then the question can be deemed non-researchable and an alternative research question may have to be posed.

The process we described above, which we consider typical of empirical research in education that involves qualitative or quantitative methods alike, is grounded in an important *assumption*: in posing a new researchable question, the researcher has sufficient understanding of the web of potentially influential factors and the possible influence that each of them might have on the phenomenon under examination so as to make informed decisions about how to handle these factors (see options a–c in the previous paragraph). This assumption, though sensible and to some extent inevitable for any empirical research study to be carried out, is often left implicit in research reports. Yet, we argue, the assumption deserves careful reflection and scrutiny in empirical research reports—both in the description of the research design and in the interpretation of the research findings about the phenomenon under examination.

The importance of doing so derives primarily from the fact that a researcher's understanding of the web of potentially influential factors for a specific research question is inevitably limited, as that understanding can only reflect the current state of research knowledge in the field about the respective topic. In other words, the researcher cannot exclude the possibility of the presence of influential factors other than those already considered or known in the literature at the time. If, however, other important factors did exist, these factors would not have been accounted for in the research design but they might have nevertheless influenced the research findings. Also, when knowledge of these other important factors becomes available, this knowledge will have to be considered by subsequent studies on the topic and, of course, it will have to be reflected in the phrasing of these studies' research questions, which in essence will be *new* researchable questions.

In this paper, we set forth the thesis that *posing new researchable questions in educational research is a dynamic process that reflects the field's growing understanding of the web of potentially influential factors surrounding the examination of a particular phenomenon.* Several important implications follow from this thesis, which we will consider later in the paper.

The Scope of the Paper

Our primary aim in this paper is to exemplify the aforementioned thesis and associated implications. We do so mainly by discussing in depth one example that derives from a major strand of mathematics education research in the area of proof. The example traces the development of researchable questions concerning researchers' efforts to investigate students' *justification schemes* (Harel & Sowder, 1998, 2007) based on students' performance on *proof construction tasks* (definitions of the terms in italics will be offered later in the paper). The area of students' justification schemes has attracted the investigation of a wealth of research questions that meet the criteria set forth by Cai, Morris, Hohensee, Hwang, Robison, Cirillo, Kramer and Hiebert (2019) that characterize significant research questions. Specifically, students' justification schemes have been connected in the literature with instructional problems that many teachers across levels of education share (see, e.g. Harel & Sowder, 2007); also researchers have strived "to understand underlying mechanisms [of these problems] and their interactions with the context" (Cai et al., 2019, p. 118).

In our discussion, we will draw partly on our own research in the area of students' justification schemes and partly on research that was conducted by others in order to illustrate the point that new researchable questions have evolved dynamically with the field's growing understanding of the web of factors that influence both students' performance on proof construction tasks and the inferences that researchers draw, or can draw, about students' justification schemes based on that performance. It is beyond the scope of this paper to trace the full development of new researchable questions in this area. One reason for this is our desire to illustrate a thesis rather than provide a comprehensive review of a particular strand of research. Another reason is our inevitably incomplete understanding of the relevant literature, which has expanded rapidly over the past few decades (for relevant partial reviews, see Harel & Sowder, 2007; Stylianides, Bieda, & Morselli, 2016; Stylianides, Stylianides, & Weber, 2017). As we will explain later in the paper, the rapid growth of research knowledge in this area has enhanced in rather complex ways the field's understanding of students' justification schemes, while at the same time it has created increased methodological challenges for researchers as they seek to achieve a defensible balance between investigating new researchable questions about students' justification schemes and being considerate of the wealth of factors that have been found to bear on students' proof constructions.

Obviously, we would not write this paper if we thought that its thesis was specific to a particular strand of research within mathematics education. Towards the end of the paper, we will briefly discuss two other strands of mathematics education research to which we argue the same thesis would apply so as to give readers a sense of the possible broader applicability of the thesis. Our limited knowledge of the evolution of researchable questions in other research strands within mathematics education and beyond does not allow us to explore the boundaries of the domain of application of the thesis. We hypothesize, though, that the same ideas would apply more broadly, including in other strands of educational research such as science education, and we invite other researchers to contribute to the exploration of this hypothesis.

Investigating Students' Justification Schemes Using Proof Construction Tasks

We will begin this section by providing some background context for research on the notion of students' justification schemes. We will then exemplify our thesis by discussing new researchable questions that emerged over time as researchers used proof construction tasks to investigate students' justification schemes. We will conduct this exemplification in two parts: first, by discussing in detail the transition from one researchable question to a new one, drawing on relevant research from the 1990s and the 2000s; second, by discussing the expansion of new researchable questions, drawing on the findings of three recent research studies in the area.

Research Context

Harel and Sowder (1998, 2007) used the notion of "justification schemes" (also referred to as "proof schemes" in some of their publications) to describe what arguments convince students and what arguments students offer to convince others, that is, what counts as a proof from the students' standpoint. Specifically, Harel and Sowder (2007) defined an individual's *justification scheme* with respect to "what constitutes ascertaining and persuading for that person" (p. 809), where *ascertaining* is defined to be "the process an individual […] employs to remove her or his […] own doubts about the truth of an assertion" (p. 808) and *persuading* is "the process an individual […] employs to remove others' doubts about the truth of an assertion" (p. 808).

Unsurprisingly, students' standards of conviction often do not align with acceptable standards of conviction in the mathematical community, and this constitutes an instructional problem that many mathematics teachers (including university instructors) face as they engage their students in proving. Indeed, students' justification schemes can take various forms, and these have been classified by Harel and Sowder (1998, 2007) under the following three general categories: (1) *externally based*, when conviction originates from some source that is external to the student, such as an authority (authoritarian justification scheme), the form of an argument (ritual justification scheme), or the treatment of symbols that does not have a particular meaning (non-referential symbolic justification scheme); (2) *empirical*, when conviction is based exclusively on the use of one or more examples (examples-based or inductive justification scheme) or on perception of one or more drawings (perceptual justification scheme); and (3) *deductive or analytic*, when conviction is based on reasoning that is concerned with the general aspects of a mathematical situation (transformational justification scheme) or on logical deduction of new results from results that are already accepted (axiomatic justification scheme).

Harel and Sowder's notion of justification schemes and respective classification framework (or modified versions thereof) have been used extensively in research studies in the area of proof. We will not discuss here other relevant classification frameworks and their relationship to Harel and Sowder's (this has been done elsewhere; see, for example Harel & Sowder, 2007, pp. 810–811; Stylianides et al., 2017, pp. 244–245). What is relevant to our purposes in this paper is that a rather large body of research has investigated students' justification schemes and has tended to use one or both of two kinds of tasks to elicit and document students' justification schemes (see,

e.g. Housman & Porter, 2003; Kanellos, Nardi, & Biza, 2018; Lee, 2016): (1) *proof construction tasks*, i.e. tasks that ask students to formulate a proof for the truth or falsity of a mathematical claim (usually a mathematical generalization); and (2) *proof evaluation tasks*, i.e. tasks that ask students to indicate whether they think given arguments for a mathematical claim (usually researcher-generated arguments of various mathematical qualities) meet the standard of proof. The operational hypothesis has been that a student's purported proof (in the first kind of task) or a student's evaluation of a given argument as meeting the standard of proof (in the second kind of task) both reflect, or indicate, the student's justification scheme.

Harel and Sowder (2007) reviewed a large number of studies of proof performance at the school and university levels, and they interpreted the findings of these studies in terms of the apparent justification schemes that could be inferred from the results. They concluded that the findings provided evidence that a pervasive justification scheme among both school and university students is the empirical justification scheme. The pervasiveness of the *empirical justification scheme* has been identified also by the Education Committee of the European Mathematical Society (2011a) as a "solid finding"[1] of mathematics education research in the area of proof:

> [C]onsiderable evidence exists that many students rely on validation by means of one or several examples to support general statements, that this phenomenon is persistent in the sense that many students continue to do so even after explicit instruction about the nature of mathematical proof, and that the phenomenon is international … (pp. 50–51)

The prominence of the empirical justification scheme among students that is described in this excerpt is essentially equivalent to many students having the *misconception that empirical arguments*[2] *are proofs*, which, as we argued elsewhere (Stylianides & Stylianides, 2009b), is a major stumbling block to students' appreciation of the need to learn about the conventional meaning of proof as a deductive argument. In our subsequent discussion, we will draw on the strand of mathematics education research that aimed to document students' justification schemes, with particular attention to whether students have the empirical justification scheme. For the purposes of our exemplification, we will focus on studies that examined students' justification schemes in the context of *proof construction tasks*, though a similar discussion could be conducted using studies that used proof evaluation tasks. We also clarify that, although we use the notion of justification schemes as an organizing structure for our

---

[1] The Education Committee of the European Mathematical Society (2011b, p. 46) defined "solid findings" as follows:
– result from trustworthy, disciplined inquiry, thus being sound and convincing in shedding light on the question(s) they set out to answer.
– are generally recogised as important contributions that have significantly influenced and/or may significantly influence the research field.
– can be applied to circumstances and/or domains beyond those involved in this particular research.
– can be summarised in a brief and comprehensible way to an interested but critical audience of non-specialists (especially mathematicians and mathematics teachers).

[2] The term *empirical argument* is generally used to describe an invalid (non-proof) argument that purports to show the truth of a statement based on verification of a proper subset of all the cases covered by the statement.

discussion, not all studies we discuss herein used this notion to frame their reported research. We judge, however, that these studies' findings (or at least the parts we are interested in) can appropriately be described in these terms and in relation to *research-able* questions that might differ from the studies' *stated* research questions.

From One Researchable Question to a New One

Some studies into students' justification schemes in the context of proof construction tasks (e.g. Healy & Hoyles, 2000; Knuth, Choppin, & Bieda, 2009; Senk, 1989), especially studies in the early stages of this research strand, addressed among others the following research question: *What are students' justification schemes as derived from students' performance on proof construction tasks?* For easy reference, we call this question RQ_1.

Essentially, the aforementioned studies asked students to prove given statements and then used the arguments produced by the students to draw conclusions about students' understanding of proof by mapping students' arguments onto Harel and Sowder's framework of justification schemes. So, for example, a student who offered an *empirical argument* in response to a task that asked the student to prove a mathematical generalization would be considered to exhibit the *empirical justification scheme*. If, on the other hand, a student produced a general mathematical argument, the student would be considered to exhibit a *deductive justification scheme*. The findings of these and other relevant studies painted a bleak picture of students' justification schemes, and it is partly on these findings that Harel and Sowder (2007) based their conclusion about the prominence of the empirical justification scheme among school and university students.

In our own research, as part of a 4-year design experiment in an undergraduate mathematics course for students who aspired to join a masters-level elementary teacher education program, we aimed among other goals to help students overcome the misconception that empirical arguments are proofs and thus help them progress beyond the empirical justification scheme (see, e.g. Stylianides & Stylianides, 2009b). However, we were surprised to observe in the early research cycles of our design experiment that a considerable number of students would persist in producing empirical arguments in response to proof construction tasks, even after we had evidence to believe that the students had understood the limitations of empirical arguments and their inadequacy to meet the standard of proof. This made us begin to uncover a potentially influential *factor* that both prior research and ourselves seemed to have overlooked up to that point in time when addressing RQ_1, namely, *students' own perceptions of whether their produced arguments actually met the standard of proof*. In particular, some students could be providing empirical arguments to proof construction tasks not because they believed that empirical arguments were proofs but because, for example, they could not come up with better arguments. In other words, instead of leaving a blank response to the question posed by the instructor, the students might have chosen to write down an erroneous response (in this case, an empirical argument) while still being fully aware of the limitations of their response (in this case, the fact that an empirical argument does not meet the standard of proof).

Our emerging appreciation of the aforementioned factor helped us realize also that our original research design was insufficient to accurately detect the possible existence of the empirical justification scheme among our students. Indeed, the design was based,

implicitly and unwarily, on an assumption that might have not been true for all students, namely, that the arguments the students produced in response to proof construction tasks were actually what the students believed constituted a proof. Accordingly, the findings we had obtained from that research design were likely inaccurate: our failure to account for the new factor might have resulted in us reporting an inflated number of students having the empirical justification scheme. Relatedly, the research question we had used in the original design (namely, RQ_1) was no longer fit for purpose as it failed to guide an investigation that would distinguish students who produced empirical arguments *and* considered those arguments as proofs (i.e. students who exhibited the empirical justification scheme) from other students who produced again empirical arguments *but* were fully aware of the limitations of their arguments (i.e. students who did not exhibit the empirical justification scheme).

To account for the new factor, we modified our research design by adding a follow-up prompt to the original prompt that asked students to prove a statement. Specifically, the follow-up prompt asked students to *evaluate* their argument constructions in response to the first prompt: "Do you believe you have actually produced a proof? Why or why not?" In the new research design, conclusions about participants' justifications schemes were drawn based on the combined consideration of their responses to the two prompts. So, for example, a student who provided an empirical argument to the first prompt (the proof construction prompt) would not be considered exhibiting the empirical justification scheme unless the student also indicated in the second prompt (the evaluation prompt) that he or she actually considered the argument to be a proof. The modification in the research design to include a "proof *evaluation* component" in addition to the "proof *construction* component" also necessitated modification of our research question. The new researchable question we posed was the following: *What are students' justification schemes as derived from students' performance on combined proof construction-evaluation tasks?* (RQ_2).

With this research design and research question, we conducted a new investigation into students' justification schemes using combined proof construction-evaluation tasks (Stylianides & Stylianides, 2009a). We found that the students who produced empirical arguments were roughly split evenly between those who were aware that their arguments did not qualify as proofs and those who believed that their arguments qualified as proofs. The same pattern was observed more generally among the students who produced non-proof arguments (of which empirical arguments is a sub-class). These findings offered evidence that students' own perceptions (evaluations) of whether their produced arguments in proof construction tasks actually met the standard of proof is an important factor for researchers to consider in investigations of students' justification schemes. The findings also helped deepen the theoretical understanding of students' justification schemes by illuminating aspects of students' proof-related behavior in proof construction tasks that tended to go undetected when students were asked to only prove given statements.

In order to offer a possible account of the phenomenon of some students providing invalid arguments to proving tasks while being aware of the limitations of their responses, we used Brousseau's (1997) notion of *didactical contract*, which refers to the system of reciprocal obligations between teachers and their students that are specific to the target knowledge. According to this system of reciprocal obligations, which may be implicit and informal, blank responses to a task that is posed by the teacher (not

necessarily a proving task) are not appropriate as they could be failing to meet the teacher's expectation to receive from the students *a* response (not necessarily a correct one), and so students tend to offer the best response they can come up with even when they are aware that their response is flawed. This account helps explain also other findings from the literature, such as those reported by Knuth et al. (2009) that students were more likely to produce empirical arguments for more difficult proving tasks (the researchers reported 81% vs. 36% empirical arguments for two tasks of different levels of difficulty). According to the researchers, "students likely had no recourse but to use examples as their means of justification" (p. 161).

The Expansion of New Researchable Questions

In this part, we will discuss briefly three recent studies that are relevant to the investigation of students' justification schemes (Dawkins & Karunakaran, 2016; Stylianides, 2019; Weber, Lew, & Mejia-Ramos, 2020), and we will consider how each of them can take RQ$_2$ to a new direction, thus giving rise to new researchable questions. We will label the new research questions as RQ$_{3a-c}$ so as to indicate their conceptual (rather than actual) evolution from RQ$_2$. We will also use RQ$_{3a-c}$ as a context in which to reflect on implications of the expansion of research knowledge in the area of students' justification schemes as well as the consequential expansion of new researchable questions in this area.

Dawkins and Karunakaran (2016) drew attention to the role of *mathematical content* (e.g. algebra, geometry, analysis), including students' mathematical meanings for mathematical content, as a potentially influential *factor* of students' proof-related behavior. This is relevant to our discussion here as we view students' justification schemes as an indicator of students' proof-related behavior. Dawkins and Karunakaran argued and illustrated the possible negative consequences of a content-generic approach to mathematics education research on students' proof-related behavior. The authors clarified that they did "not intend to deny the validity or value of prior research framed in a content-independent manner […] but rather seek to sensitize the community to possible blind spots induced by common lenses applied to research data and to endorse a research agenda focused on the interplay between proving and particular mathematical content" (p. 65). Dawkins and Karunakaran's message implies a need for the role of mathematical content to be considered in examinations of students' justification schemes, including in the statement of research questions. This message leads, then, to an extension of RQ$_2$ to the following new set of researchable questions: *In the particular content area of algebra/geometry/analysis/[…], what are students' justification schemes as derived from students' performance on combined proof construction-evaluation tasks?* (RQ$_{3a}$). Comparisons of students' proof constructions and evaluations of their constructions in different content areas can offer useful insights into the extent to which students' justification schemes are content-specific or content-generic.

Moving on to the second study we will consider in this part, Stylianides (2019) drew attention to another *factor* that received limited attention by research on students' proof-related behavior, namely, *the mode of representation (written* versus *oral) used by students as they communicate their proof constructions.* In a way similar to Dawkins and Karunakaran (2016) but with a focus on the role of argument representation, Stylianides (2019) argued and illustrated the possible negative consequences of a

representation-generic examination of students' proof constructions. Specifically, in the context of actual classroom practice, Stylianides compared the written arguments that secondary students produced and perceived to be proofs with the oral arguments that the same students presented in front of the class for the same claims, and he found that the oral arguments were more likely than the written arguments to meet the standard of proof. In discussing the implications of these findings, Stylianides noted the following:

> [P]rior research on secondary students' argument constructions tended to use survey methods and only consider students presenting their perceived proofs in written form. Based on students' written arguments, this research has painted a bleak picture of secondary students' ability to construct arguments that meet the standard of proof. The findings reported in this article suggest that the limited use of observational methods and the lack of consideration of students presenting their perceived proofs orally—in tandem with students' written proofs for the same claims—is a serious threat to the validity of research findings in this area. Indeed, if a study had analysed students' written arguments only (as in survey research), it would have reported a less favourable picture of the potential of students' constructed proofs than another study that would focus only on students' oral arguments (as in observational research). Also, by considering only one mode of representation and ignoring the other, each study individually would have reported an incomplete picture of students' constructed proofs, for apparently it matters whether students present their perceived proofs orally or in writing. (p. 177)

Stylianides' findings support another extension of RQ_2 as follows: *What are students' justification schemes as derived from students' performance on combined proof construction-evaluation tasks whereby students' perceived proofs are communicated in written/oral/combined written-and-oral form?* (RQ_{3b}).

Weber et al. (2020) further expanded work in this area by proposing a new framework for investigating and explaining students' proof-related behavior. The framework introduces additional *factors* that potentially influence students' proof constructions and the inferences that researchers might draw from them about students' justification schemes. Specifically, Weber et al. adapted constructs from "expectancy value theory" (e.g. Eccles & Wigfield, 2002) to investigate and illustrate empirically the following claim: whether a student will seek a proof or settle for a non-proof argument (including an empirical argument) depends partly on the *value* the student places on knowing the veracity of a mathematical statement, the *cost* in terms of time and effort that is required for the search for a proof, and the student's perceived *likelihood of success* of being able to find a proof. Consider, for example, a student who offered an empirical argument in response to a proof construction task. This student might believe that the empirical argument bestowed certainty in the truth of the generalization and thus consider the argument to be a proof, in which case the student can be said to exhibit the empirical justification scheme. However, Weber et al. argued that there are at least three other possibilities for which the student might have offered the empirical argument: (1) the student might not be interested in being certain about the truth of the generalization and thus settle for the first argument (empirical) that he or she produced (an issue of *value*); (2) the student might consider searching for a proof to be an

unpleasant endeavor and thus settle for an empirical argument that presumably requires less effort to produce than a proof (an issue of *cost*); or (3) the student might settle for the empirical argument because he or she believes the construction of a proof is beyond his or her capability (an issue of *likelihood of success*). These alternative possibilities of the student's proof-related behavior challenge the certainty of the conclusion that a student's empirical argument is evidence of the student exhibiting the empirical justification scheme and thus point to the need for research into students' justification schemes to consider the three factors in the expectancy value model. In other words, this model supports another extension of RQ_2 as follows: *What are students' justification schemes as derived from students' performance on combined proof construction-evaluation tasks and based on students' considerations of value, cost, and likelihood of success in their produced arguments?* (RQ_{3c}).

The three studies in the area of students' justification schemes that we discussed in this part of the paper (Dawkins & Karunakaran, 2016; Stylianides, 2019; Weber et al., 2020) illustrate the idea that the conceptual terrain in this research area becomes more complex with advancements in research knowledge about the factors that can influence students' performance on proof construction tasks. Also, as the field's knowledge of new potentially influential factors grows, there is a consequential expansion of new researchable questions that can guide researchers' investigations towards a more refined, and presumably more accurate, body of research knowledge about students' justification schemes. For example, findings of investigations of new researchable questions RQ_{3a} and RQ_{3b} can contribute, respectively, to the development of content-specific (Dawkins & Karunakaran, 2016) and representation-specific (Stylianides, 2019) portraits of students' justification schemes, which in turn might help point out important trends that went undetected by prior research. At the same time, however, the field's research knowledge about students' justification schemes can become more fragmented: As research reports of content- and representation-generic findings give place to reports of content- and representation-specific findings, it is less justifiable for researchers to cluster together or even compare findings across studies that did not account for these factors. Although this may be an unwelcome development from a policy standpoint where generalized descriptions of phenomena can help guide policy decisions about instructional practice, things can be different from a research standpoint. Dawkins and Karunakaran (2016) hypothesized that "our field's implicit invitation to overgeneralize empirical findings are partly to blame for the confusing and seemingly contradictory claims available in the literature on proof" (p. 73).

The field's growing knowledge of additional potentially influential factors renders the methodological landscape complex too. Clearly, it will be methodologically challenging to design a new research study to document students' justification schemes in the context of proof construction tasks that will account for the whole range of factors reflected in $RQ_{3a–c}$: the mathematical content where the tasks are embedded (RQ_{3a}: Dawkins & Karunakaran, 2016), the mode of representation with which students' arguments are communicated (RQ_{3b}, Stylianides, 2019), and students' considerations of value, cost, and likelihood of success in constructing their arguments (RQ_{3c}, Weber et al., 2020). One can appreciate further the challenges of designing such a study by being mindful of the fact that the potentially influential factors worth considering might not be limited to those that we discussed herein for the purposes of exemplification.

We clarify that we are not suggesting that all new research studies aiming to investigate students' justification schemes should account for all of the potentially influential factors discussed herein or others. We share Dawkins and Karunakaran's (2016) view that "students' mathematical reasoning is an incredibly multi-faceted and complex forum for investigation" and that "[n]o single study can account for all of the dimensions of variation at play" (p. 73). As we pointed out in the opening paragraph of this paper, when posing a new researchable question, a researcher needs to make decisions about how to deal with the whole web of factors that can possibly influence the phenomenon under examination in a careful, transparent, and justifiable way, drawing on relevant theoretical or conceptual frameworks and associated empirical findings. But how can researchers deal, in practical terms, with the fundamental challenge of curtailing and managing the proliferation of factors that can influence the phenomenon under examination? We will consider this question in the next section as it is broad and does not relate specifically to the notion of justification schemes.

Concluding Remarks

In this paper, we argued that posing new researchable questions in educational research is a dynamic process that reflects the field's growing understanding of the web of potentially influential factors surrounding the examination of a particular phenomenon of interest. We illustrated this thesis by drawing on a strand of mathematics education research related to students' *justification schemes* (Harel & Sowder, 1998, 2007) that has evolved rapidly during the past few decades (see, e.g. Harel & Sowder, 2007; Stylianides et al., 2016, 2017) and thus offered a good context for exemplification of the thesis. In what follows we will first reflect on the boundaries of the domain of application of our thesis, and then we will consider three main implications of the thesis.

Boundaries of the Domain of Application of the Thesis

We cannot accurately determine the boundaries of the domain of application of the thesis, because we lack a thorough knowledge of the evolution of researchable questions across strands of educational research. However, our (limited) knowledge of several other research strands within mathematics education gives us confidence that the domain of application of the thesis is broad and goes beyond that of justification schemes. Next, we will briefly consider two other examples to illustrate this point.

The first example relates to the following research question: "What is the potential of mathematical problems, with or without a connection to reality, to trigger students' task-specific interest in problem solving?" In describing this example, we draw primarily on work by Schukajlow and colleagues (Rellensmann & Schukajlow, 2017; Schukajlow, Leiss, Pekrun, Blum, Müller, & Messner, 2012) who distinguished between two broad categories of problems: those with a connection to reality, called "real-world problems," and those without such a connection to reality, called "intra-mathematical problems." Although there are sub-categories within each of these categories of problems, real-world problems are often used in curriculum materials for the purpose of triggering student interest (Meyer, Dekker, & Querelle, 2001). However, there is a lack of robust empirical evidence linking any of these problem categories to a higher level of

student interest. For example, Schukajlow et al. (2012) found no difference in students' interest in real-world and intra-mathematical problems. In a follow-up study, Rellensmann and Schukajlow (2017) investigated the above research question by controlling for *task difficulty*, which was a factor that prior research including Schukajlow et al. (2012) had tended to overlook. Controlling for task difficulty resulted in the reversal of Rellensmann and Schukajlow's (2017) original expectations, for they found that "intra-mathematical problems [rather than real-world problems] are better suited for capturing students' interest" (p. 375). According to the authors, "[t]hese findings underpin the idea that task difficulty is a factor that should be taken into account in research on task-specific interest (Renninger, 1998) because the difference in students' interest in solving problems with and without a connection to reality is hidden if the confounding effect of task difficulty is not controlled for" (p. 375).

The second example relates to the following research question that was posed in the 1970s: "What is the effect of teachers' mathematics coursework on students' mathematics achievement?" Begle's (1979) meta-analysis of studies conducted between 1960 and 1976 (cited in Ball, Lubienski, & Mewborn, 2001) showed that teachers' mathematics coursework produced positive main effects on students' achievement in only 10% of the cases and, more surprisingly, negative main effects in 8%. These studies used teachers' coursework in mathematics as a proxy for teachers' mathematics content knowledge. Given that teachers' mathematics content knowledge was deeply thought to make a difference in students' achievement, the aforementioned findings (especially the negative effects) puzzled researchers over the role of teachers' content knowledge in teaching. Subsequent theoretical advances in the field elaborated the notion of teachers' content knowledge and its various sub-components (e.g. Ball, Thames, & Phelps, 2008; Rowland, Turner, Thwaites, & Huckstep, 2009; Shulman, 1986), making clear that teachers' coursework is a poor proxy for the multi-faceted notion of teachers' content knowledge, which includes also components such as teachers' understanding of students' thinking and ways of representing the subject matter. These theoretical advances gave rise to new notions that are used to describe teachers' knowledge, like *teachers' mathematical knowledge for teaching* (Ball et al., 2001, 2008), and they led to the posing of new researchable questions such as the following: "What is the effect of teachers' mathematical knowledge for teaching on students' mathematics achievement?" (Hill, Rowan, & Ball, 2005). Hill et al. (2005) found that teachers' mathematical knowledge for teaching has a positive effect on students' mathematics achievement, a finding that marks a major advancement in research knowledge since Begle's (1979) time. This is not to criticize Begle who posed a research question that made sense at the time, but to illustrate the point that an increased understanding of the web of potentially influential factors led to new researchable questions.

Implications of the Thesis

Besides serving the purpose of deepening theoretical understanding about how new researchable questions are, or can be, generated in educational research, our discussion also identified and illustrated three important implications of our thesis. Next, we will summarize and elaborate as appropriate on these inter-related implications.

The first implication is that, as new potentially influential factors are identified, findings from past research that had not accounted for those factors might prove to be

insufficient or be challenged. This implication is illustrated by all three examples we discussed in this paper. Clearly, this is not a criticism of past research as the findings of those studies are judged retrospectively, but the identification of new potentially influential factors does nevertheless complexify the conceptual landscape surrounding the phenomenon of interest.

The second implication concerns the methodological challenges that arise for researchers as they navigate this complex conceptual landscape and seek to design new studies that pay due regard to research advances about all relevant and potentially influential factors. Of course it is still possible for new studies to focus on specific factors pertaining to the phenomenon of interest, say the justification schemes exhibited by students as they engage with proving tasks involving visual or non-visual reasoning (cf. Alcock & Simpson, 2004, 2005; Hadas, Hershkowitz, & Schwarz, 2000), while accounting for or considering in a meaningful way other known relevant factors, such as by limiting the research scope (and conclusions) in a particular content area (e.g. geometry). The broader question that arises at this point, though, is the one that we posed earlier in the paper and relates to the dialectical nature of the web of factors known to influence a phenomenon and investigations of that phenomenon: How can researchers deal, in practical terms, with the fundamental challenge of curtailing and managing the proliferation of factors that can influence the phenomenon under examination? This is a particularly hard question to address and possibly deserves a paper of its own. We raise it here explicitly to offer some initial thoughts about it, drawing primarily on Clement's (2000) discussion of "explanatory models," and to invite other researchers to unpack it and elaborate on possible ways to address it.[3] As we will explain, Clement's discussion of explanatory models can be used to offer researchers theoretically principled ways to justify their focus on particular factors while ignoring others.

According to Clement (2000), *explanatory models* "are not merely condensed summaries of empirical observations but, rather, are inventions that contribute new mechanisms and concepts that are part of the scientist's view of the world and that are not 'given' in the data" (p. 549). In other words, Clement views such models as providing an explanatory description why the phenomenon of interest occurred and give satisfying explanations for "patterns of observable behavior." Using again the notion of students' justification schemes as an example and applying Clement's terminology, a student's empirical justification would be an "observable behavior" while students' propensity to offer empirical justifications would be a "pattern of observable behavior." The factors that we discussed in the previous section of the paper related to the mathematical content where the proving tasks are embedded (cf. Dawkins & Karunakaran, 2016), the mode of representation with which students' arguments are communicated (cf. Stylianides, 2019), and students' considerations of value, cost, and likelihood of success in constructing their arguments (cf. Weber et al., 2020) are all examples of factors that may well be part of an explanatory model of the aforementioned pattern of observable behavior. Yet developing an explanatory model is challenging, especially when it pertains to higher order thinking skills such as students' justification schemes that are complex and prone to the influence of a multitude of factors. Another challenge in designing an explanatory model is finding an appropriate

---

[3] We are grateful to an anonymous reviewer who brought Clement's (2000) work to our attention.

grain size for it. As Clement points out, there are many virtues of an explanatory model that go beyond accuracy and that there are trade-offs as one cannot simply act to maximize all the characteristics of a good model. Parsimonious models are often superior to complicated models that are slightly more accurate.

Accordingly, a good parsimonious model can offer a way to address the question we raised above about the dialectical nature of the web of factors known to influence a phenomenon and investigations of that phenomenon, by allowing researchers, in a theoretically principled way, to cleave out the list of factors they want to consider in their study while ignoring others. Until such a model is developed, though, it is important that reports of research include a detailed description of the context of the studies as well as the factors that were considered in the studies and other known factors that might have influenced the phenomenon but were not considered. This detailed description of the context is particularly important because, if new factors that influence the phenomenon of interest are discovered in the future, researchers can be in a position to determine whether those factors had any bearing on the findings of past studies and the extent to which the findings of those studies are still useful for comparative purposes or applicable in specific contexts.

The third implication of our thesis in this paper builds on the previous two and concerns the evolving nature of research knowledge as findings from studies on new researchable questions become available. For example, early research on students' justification schemes could defensibly draw general conclusions like "many students exhibit the empirical justification scheme." However, the more we (as a field) learn about potentially influential factors, the more we view such general conclusions with caution. In particular, the aforementioned conclusion would now have to be qualified in terms of the mathematical content where students exhibited certain schemes (cf. Dawkins & Karunakaran, 2016) or in terms of the mode of representation that students used to communicate their arguments (cf. Stylianides, 2019). Thus, as a wider range of potentially influential factors get discovered and considered, research knowledge becomes not only more refined and presumably more accurate, but possibly more fragmented too. On this basis, one may be confronted with this paradox: As we (as a field) deepen our understanding of the complex network of factors that have a bearing on the phenomenon of interest, the further away we get from being in a position to draw defensible general conclusions about the phenomenon. This paradox, which may not be specific to students' justification schemes or the other two examples we discussed in this paper, has policy ramifications, especially with regard to assessment. For example, assessing students' justification schemes cannot be a matter of evaluating students' performance on proving tasks from a single content area (popularly geometry) or in written form only.

References

Alcock, L., & Simpson, A. (2004). Convergence of sequences and series: Interactions between visual reasoning and the learner's beliefs about their own role. *Educational Studies in Mathematics, 57*, 1–32.

Alcock, L., & Simpson, A. (2005). Convergence of sequences and series 2: Interactions between non-visual reasoning and the learner's beliefs about their own role. *Educational Studies in Mathematics, 58*, 77–110.

Ball, D. L., Lubienski, S., & Mewborn, D. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 433–456). New York, NY: Macmillan.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389–407.

Begle, E. G. (1979). *Critical variables in mathematics education: Findings from a survey of the empirical literature*. Washington, DC: Mathematical Association of America and National Council of Teachers of Mathematics.

Brousseau, G. (1997). *Theory of didactical situations in mathematics: Didactique des mathématiques 1970–1990* (N. Balacheff, M. Cooper, R. Sutherland, & V. Warfield, Trans. & Eds.). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., Cirillo, M., Hiebert, J. (2019). Posing significant research questions. *Journal for Research in Mathematics Education, 50*(2), 114–120.

Clement, J. (2000). Analysis of clinical interviews: Foundations and model viability. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 547–589). Mahwah, NJ: Lawrence Erlbaum.

Dawkins, P. C., & Karunakaran, S. S. (2016). Why research on proof-oriented mathematical behavior should attend to the role of particular mathematical content. *The Journal of Mathematical Behavior, 44*, 65–75.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*, 109–132.

Education Committee of the European Mathematical Society. (2011a). Do theorems admit exceptions? Solid findings in mathematics education on empirical proof schemes. *EMS Newsletter, 82*, 50–53.

Education Committee of the European Mathematical Society. (2011b). "Solid findings" in mathematics education. *EMS Newsletter, 81*, 46–48.

Hadas, N., Hershkowitz, R., & Schwarz, B. B. (2000). The role of contradiction and uncertainty in promoting the need to prove in dynamic geometry environments. *Educational Studies in Mathematics, 44*, 127–150.

Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. In A. Schoenfeld, J. Kaput, & E. Dubinsky (Eds.), *Research in collegiate mathematics education III* (pp. 234–283). Providence, RI: American Mathematical Society.

Harel, G., & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 805–842). Greenwich, CT: Information Age Publishing.

Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education, 31*, 396–428.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406.

Housman, D., & Porter, M. (2003). Proof schemes and learning strategies of above-average mathematics students. *Educational Studies in Mathematics, 53*(2), 139–158.

Kanellos, I., Nardi, E., & Biza, I. (2018). Proof schemes combined: Mapping secondary students' multi-faceted and evolving first encounters with mathematical proof. *Mathematical Thinking and Learning, 20*(4), 277–294.

Knuth, E., Choppin, J., & Bieda, K. (2009). Middle school students' production of mathematical justifications. In D. Stylianou, M. Blanton, & E. Knuth (Eds.), *Teaching and learning proof across the grades: A K–16 perspective* (pp. 153–170). New York, NY: Routledge.

Lee, K. (2016). Students' proof schemes for mathematical proving and disproving of propositions. *The Journal of Mathematical Behavior, 41*, 26–44.

Meyer, M., Dekker, T., & Querelle, N. (2001). Context in mathematics curricula. *Mathematics Teaching in the Middle School, 6*(9), 522–527.

Rellensmann, J., & Schukajlow, S. (2017). Does students' interest in a mathematical problem depend on the problem's connection to reality? An analysis of students' interest and pre-service teachers' judgments of students' interest in problem with and without a connection to reality. *ZDM Mathematics Education, 49*, 367–378.

Renninger, K. A. (1998). What are the roles of individual interest, task difficulty, and gender in student comprehension? In L. Hoffmann, J. Baumert, A. Krapp, & K. A. Renninger (Eds.), *Interest and learning: Proceedings of the Seeon Conference on interest and gender* (pp. 228–238). Kiel, Germany: IPN.

Rowland, T., Turner, F., Thwaites, A., & Huckstep, P. (2009). *Developing primary mathematics teaching: Reflecting on practice with the knowledge quartet*. London, England: Sage.

Schukajlow, S., Leiss, D., Pekrun, R., Blum, W., Müller, M., & Messner, R. (2012). Teaching methods for modelling problems and students' task-specific enjoyment, value, interest and self-efficacy expectations. *Educational Studies in Mathematics, 79*, 215–237.

Senk, S. L. (1989). van Hiele levels and achievement in writing geometry proofs. *Journal for Research in Mathematics Education, 20*(3), 309–321.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14.

Stylianides, A. J. (2019). Secondary students' proof constructions in mathematics: The role of written vs. oral mode of argument representation. *Review of Education, 7*(1), 156–182.

Stylianides, A. J., Bieda, K. N., & Morselli, F. (2016). Proof and argumentation in mathematics education research. In A. Gutiérrez, G. C. Leder, & P. Boero (Eds.), *The second handbook of research on the psychology of mathematics education*(pp. 315–351). Rotterdam, The Netherlands: Sense Publishers.

Stylianides, A. J., & Stylianides, G. J. (2009a). Proof constructions and evaluations. *Educational Studies in Mathematics, 72*(2), 237–253.

Stylianides, G. J., & Stylianides, A. J. (2009b). Facilitating the transition from empirical arguments to proof. *Journal for Research in Mathematics Education, 40*, 314–352.

Stylianides, G. J., Stylianides, A. J., & Weber, K. (2017). Research on the teaching and learning of proof: taking stock and moving forward. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 237–266). Reston, VA: National Council of Teachers of Mathematics.

Weber, K., Lew, K., & Mejia-Ramos, J. P. (2020). Using expectancy value theory to account for individuals' mathematical justifications. *Cognition & Instruction, 38*(1), 27–56.

Affiliations

Gabriel J. Stylianides[1] · Andreas J. Stylianides[2]

[1]   Department of Education, University of Oxford, 15 Norham Gardens, Oxford OX2 6PY, UK

[2]   Faculty of Education, University of Cambridge, 184 Hills Road, Cambridge CB2 8PQ, UK