



# There is an Evidence Crisis in Science Educational Policy

Lin Zhang<sup>1</sup> · Paul A. Kirschner<sup>2,3</sup> · William W. Cobern<sup>4</sup> · John Sweller<sup>5</sup>

Accepted: 25 October 2021 / Published online: 6 November 2021  
© The Author(s) 2023, corrected publication 2024

## Abstract

There is a considerable gap between many of the findings from educational psychology research and educational practice. This gap is especially notable in the field of science education. In this article, the implications of three categories of research and their findings for science educational policy in the USA and other jurisdictions were reviewed. We indicate that a particular category of research that we call “Program-Based Studies,” has dominated the formulation of educational standards while a large number of critical findings from randomized, controlled studies and correlational studies that overwhelmingly show minimal support for the suggested policy have been marked as irrelevant and excluded. The current blanket-emphasis on program-based studies at the expense of the other types of research is misplaced. Educational standards should represent a balanced view of the available data including findings from controlled and correlational studies. Finally, we indicate how these different forms of research might inform each other and provide coherent and consistent implications for educational procedures.

**Keywords** Methodology · Educational practice · Educational policy · Controlled studies · Correlational studies · Program-based studies · Science education

## Introduction

Educational psychology has the potential to provide findings to inform a variety of educational teaching and learning decisions and practices. That potential is frequently but not always realized. We identify a disconnect between educational suggestions made by extensive bodies of educational psychology research findings and common educational practices promoting inquiry and exploration as a pedagogy. A lack of discussion on various categories of research findings and their different implications for educational practice has contributed to this

---

✉ Lin Zhang  
lzhang@providence.edu

Extended author information available on the last page of the article

disconnect, leading in turn, to an evidence crisis with adverse consequences for the formulation of educational policy. We unfold this evidence crisis in science education as it is a heavily researched field with a broad cross sample of research methodologies. More importantly, the exploration-based pedagogy, frequently called “inquiry,” “discovery,” “problem-based,” or “investigations,” has been prominently reflected in science education practice and policy for decades in the USA and internationally.

The emphasis on incorporating scientific investigation in science curricula has been a global phenomenon and is commonplace. A few examples are as follows: In Australia, ACARA (Australian Curriculum Assessment and Reporting Authority, 2018) revised and released its Year 10 Australian Curriculum to emphasize students’ involvement in scientific inquiry methods to conduct investigations. More recently, new draft versions of policies in all school curricula areas were released for comment and feedback and are currently under revision after considerable public controversy over the emphasis on inquiry-based learning. In the Netherlands, curriculum reforms, such as Our Education2032 (Platform Onderwijs, 2032, 2016) and, more recently, curriculum.now (Coördinatiegroep curriculum.nu, 2019) have focused primarily on emphasizing inquiry learning. Similarly, in Canada, a curriculum was restructured around an inquiry process model to encourage students to ask and pursue questions through scientific explorations and investigations (British Columbia Ministry of Education, 2015). As part of the global community, science education in the USA has played an important role in advocating the integration of investigations into science teaching and in leading curriculum reforms worldwide. This emphasis is disconnected from some very common forms of research, namely randomized controlled studies and correlational studies. Such a disconnect is inevitably noticeable in the field of science education and has led to consequences in the formation of educational policy world-wide. We will focus the rest of our discussion on the USA.

In the early 1960s, American educators responded to the Soviet Union’s Sputnik launch—rarely mentioned today (Kirschner, 2000)—embracing the assumption that the best way to educate young “scientists” was to use the epistemology of the expert scientists (Kirschner, 1992). Investigations, usually conducted in science laboratories but also used as a more general instructional/pedagogical classroom technique, were introduced to provide learners with opportunities to come up with their own understanding of the discipline through experimentation. For example, scientific inquiry was promoted in the 1996 National Science Education Standards (NSES) as an instructional approach and science teaching shifted from directly providing students with established content to having them go through a scientific inquiry process, including formulating research questions, making observations, collecting and analyzing data, and drawing conclusions (National Research Council, 1996).

In the years that followed, some categories of evidence supporting this approach became available, including attempts to improve students’ understandings of science concepts, develop ownership of the knowledge, foster positive attitudes toward science, and promote practical skills in authentic settings (Barton & Tan, 2010; Edelson & Reiser, 2006; Geier et al., 2008; Mistler-Jackson & Songer, 2000; Shaffer, 2004; Songer et al., 2002, 2003; Williams & Linn, 2002). Much of this evidence was

from field-initiated projects and programs grounded in a category of research we have called ‘Program-Based Studies’ (see below).

The *Next Generation Science Standards* (NGSS) (NGSS Lead States, 2013) is the most recent American effort at developing national science standards, and in these standards “inquiry” as a process has been replaced by “scientific practice.” Scientific investigation continues to be emphasized, but characterized as scientists practicing the science profession to discover the unknown. It is clear that the newly developed standards intend to further integrate the development of scientific ideas with the engagement of scientific practice, as articulated in the Framework for the NGSS (National Research Council, 2012, p. x in Foreword Section), and argue that both are essential to science learning. These arguments, as in the 1996 NRC NSES, emphasize that science learning goes beyond science content and that there is a much broader set of outcomes for science education in this new century, from strengthening a deeper understanding of science, developing a greater interest in and appreciation of science, to promoting higher scientific skills and competency to enter careers, to foster social communications, and to deal with this ever-changing world. It is also worth mentioning that even a quarter of a century ago, Hodson (1996) summarized and debated this same set of arguments used by educators for developing a broader set of science education learning outcomes. The issues are perennial.

Indeed, the rapidly evolving world requires education to carefully provide and prepare future citizens with the knowledge and skills that they will need. Educational standards need to set clear expectations and guide educators with a detailed plan to achieve them. We have no objections to standards documents that provide expected learning outcomes. We are concerned however, that while it is appropriate for standards to provide outcome expectations, the current standards also promote certain instructional approaches for achieving these expectations and that the suggested approaches are based almost exclusively on only one category of research, namely program-based studies, neglecting other research categories.

For example, as a way to integrate three dimensions (science and engineering practices, science disciplinary core ideas, and crosscutting concepts), the standards often suggest that students “plan and conduct investigations to determine...” (NGSS Lead States, 2013, 1-PS4-3. & 2-LS2-1.) where the core outcomes are treated as a likely or even inevitable product of the investigation process. The Framework that is used to philosophically and empirically guide the development of the NGSS states that “Core ideas in the framework are specified not as explanations to be consumed by learners...The practices include several methods for generating and using evidence to develop, refine, and apply scientific explanations to construct accounts of scientific phenomena...The expectation is that students generate and interpret evidence and develop explanations of the natural world through sustained investigations” (National Research Council, 2012, pp. 254–255). There is no suggestion provided in those documents that these goals might be better reached, at least under some circumstances, by procedures other than investigations, such as “explanations to be consumed by learners.”

We are concerned that while NGSS provides appropriately clear learning outcome expectations, it is also committed to teaching science core ideas by having students achieve them primarily, if not exclusively, by going through investigation steps

as an exploration process where the answers to the investigations are not presented to students, but obtained during the investigation process. We argue that although the terms used to characterize the scientific process have been changed over the years from “inquiry” to “scientific practice,” their emphasis on teaching science through investigations has remained and their characterization of the investigation procedure as an instructional/pedagogical approach is the common ground shared between the 1996 and 2013 standards.

Science education scholars such as García-Carmona (2020) noted that these two sets of standards are almost identical in terms of their promoting scientific process in science instruction and argued that the elements used to characterize the activities undertaken by scientists were not substantially different in the NGSS from the previously promoted inquiry approach of the 1996 standards. We are concerned that the same rationale is reinforced to justify a need for this newly named but pedagogically identical approach to science teaching while psychologically oriented studies that often isolate factors in controlled conditions to identify causal effects are marked as “simplistic,” even “mistaken,” and are systematically excluded (National Research Council, 2012, p. 253). We emphasize that a series of differently named instructional strategies, including problem-based learning, discovery learning, inquiry-based learning, exploration-based investigation, all withhold some direct, explicit instruction and replace it with different forms of problem solving and that they have dominated the way that scientific investigations are characterized in science instruction. As a result, there is a limited view of how scientific investigations should be integrated into science teaching to develop scientific ideas and skills. The field has been heading in an inquiry learning/investigations direction over decades while ignoring a considerable body of psychologically oriented findings that speak directly to the pedagogical principles shared among all these differently named strategies.

We never should have reached the current point, as accumulated evidence from controlled studies, on which the field of educational psychology relies heavily, has found minimal support for teaching science through exploration-based investigations. For example, controlled studies compared the aforementioned inquiry- or exploration-based investigation approach to science teaching with various forms of explicit instruction, such as simply providing students with the desired information and having students read it from texts or watch demonstrations. Overwhelmingly, these studies supported the suggestion that students demonstrated greater learning of the content from the various forms of explicit instruction (Ashman et al., 2020; Hsu et al., 2015; Klahr & Nigam, 2004; Matlen & Klahr, 2013; Renken & Nunez, 2010; Stull & Mayer, 2007; Zhang, 2018, 2019). It should be noted that many of these studies were not carried out in a psychological laboratory, but used real classrooms with students studying relevant curriculum materials.

Supporting the findings from controlled studies, an entirely different category of research, correlational studies, have used large data sets, such as TIMSS (Trends in International Mathematics and Science Study) and PISA (Program for International Student Assessment), to examine the effectiveness of an investigation approach to science teaching over large populations. Consistently, these studies across years and countries found that the more students were involved in exploration-based investigations, the lower was their performance (Aditomo &

Klieme, 2020; Areepattamannil, 2012; Cairns & Areepattamannil, 2019; Forbes et al., 2020; Gao et al., 2017; Jerrim et al., 2019; Kaya & Rice, 2010; Lavonen & Laaksonen, 2009; Liou, 2020; McConney et al., 2014; Zhang & Li, 2019). Reviewing this accumulated evidence, we are left to ponder how research evidence has been selected to develop suggestions for educational policies, such as the NGSS.

The divergence of views has been debated by many scholars (Hmelo-Silver et al., 2007; Kuhn, 2007; Lee & Anderson, 2013; Matlen & Klahr, 2013; Schmidt et al., 2007; Sweller et al., 2007; Tobias & Duffy, 2009; Zhang, 2016). The NGSS was developed in 2013, several years after the initial papers from these debates were published, but nevertheless, promoting the same incorporation of investigations in science teaching and addressing none of the issues and concerns raised in the debate. Currently, a new report has been published with a call to action (National Academies of Sciences, Engineering, and Medicine, 2021) which addresses critical issues in science education but employs the same way of characterizing investigations in science instruction as in the past, despite all prior efforts. We acknowledge the complexity of the issues but are concerned that policy documents are ignoring findings unfavorable to inquiry-based approaches and that they are automatically assumed to be irrelevant. As the field continues to grow, inevitably more examples of findings incompatible with current procedures will emerge. Those incompatibilities should be discussed and if possible, resolved.

In the following, we unfold the disconnections between different forms of research and their suggestions for science education. To be clear and specific about the issues we intend to present here, we focus our debate on contradictory suggestions made for enhancing students' learning of science concepts and procedures. We believe that science conceptual and procedural understandings are indispensable to the high-level learning of any science topics. As noted earlier, we have no objections to the various science learning outcomes that are strived for. Nor do we reject any data indicating that investigation activities can be included to achieve different learning goals. We argue however, that:

1. The development of students' science conceptual knowledge is not best obtained by having students go through exploration-based investigation activities.
2. Although we hold that scientific procedures are an essential part of science education, we do not believe that investigative skills and methods in specific science fields emerge automatically as students engage in such investigation activities. Rather, they need to be explicitly and directly taught and then sufficiently practiced in guided or open situations. We are also aware that the current standards tend to emphasize the development of a generic set of inquiry/problem-solving skills covering several science subject fields. The expectation is that once students acquire these so-called general problem-solving skills in their early education they will be able to perform better in specific fields when they launch their careers in future. While the acquisition of such skills is debatable, there can be no doubt that for students to be able to successfully carry out scientific investigations, they need to acquire conceptual and procedural content.

3. The development of other related science learning goals during investigation activities, such as attitudes toward science, should not be at the cost of students' learning of science concepts and procedures.

We suggest that the disjoint between policy documents and research evidence is exacerbated by the tendency to ignore categories of research that do not provide the favored research outcomes that support teaching science through inquiry and investigations. We also indicate how different forms of research might inform and strengthen each other during interdisciplinary collaborations to provide coherent and consistent implications for educational practice. Finally, we discuss the barriers to arriving at research-based conclusions.

## **Disconnections Between Various Forms of Research and Their Different Implications**

In this section, we describe three different categories of research, namely program-based studies, randomized controlled studies, and correlational studies. We point out that program-based studies have yielded different results from the other two categories.

### **Program-Based Studies**

Integrating scientific investigation steps into science teaching has been the focus of science educational reforms for decades. As we have already noted, the 1996 NRC NSES promoted an inquiry-based approach to science teaching and claimed that “engaging students in inquiry helps students develop understanding of scientific concepts, an appreciation of ‘how we know’ what we know in science, understanding of the nature of science, skills necessary to become independent inquirers about the natural world, and the dispositions to use the skills, abilities, and attitudes associated with science” (National Research Council, 1996, p. 105). A major problem with the NSES is that the references it used to support these claims and standards were more theoretical ideas packaged in conceptual articles rather than empirical evidence.

Following the 1996 NRC science educational standards, there were nation-wide, sweeping curriculum reform efforts. Many centers and programs were funded and built to promote inquiry-based science teaching and program-based studies were prevalent. We highlight this type of study as a key category for discussion because such work has been representative, indeed dominant, in driving educational practice.

This type of study does use quantitative data to measure the effectiveness of a program. In addition, qualitative data are also incorporated to provide detailed descriptions of students' learning experiences in the program. Among many of these studies, a pre-post research design was often used. For example, the Kids as Global Scientists (KGS) program was funded and built using scientific inquiry as its overarching instructional approach. By comparing students' pre-post

performance, the research team found that participating students significantly improved their understandings of science content and inquiry skills (Songer et al., 2002). Qualitative data, like classroom observations and teacher interviews, were also collected to provide examples of classroom teaching. No doubt, these researchers built a rigorous science learning program and brought quality learning experiences to students. Nevertheless, the positive results obtained using this research design without, for example, a proper control group, may only reflect the effectiveness of the program in terms of its capability in achieving desired student learning outcomes. It does not provide evidence for the effectiveness of inquiry-based instructional approaches over other methods due to its lack of comparisons with a control group or other teaching approaches. The authors themselves stated, “this study was not intended to demonstrate the effectiveness of the KGS intervention on student learning as compared to other interventions or traditional teaching methods” (Songer et al., 2002, p. 137). We share that view.

Program-based studies frequently also use research designs that compared student learning outcomes in the treatment group with those receiving “business as usual” as a way to evaluate the impact of a program. These programs often incorporate many instructional elements, from various forms of technology support to intensive teachers’ professional development workshops. Thus, it is questionable that students’ learning gains can be attributed only to the use of an inquiry-based approach, rather than other factors these programs also implement. For example, the Center for Learning Technologies in Urban Schools (LeTUS) is another inquiry-based learning program built during that period. The research team compared LeTUS participants’ learning outcome with those who did not participate and found that students enrolled in LeTUS inquiry-based science curricula performed significantly better (Geier et al., 2008). Nevertheless, because this program built in various instructional elements (e.g., the technologies used, the teacher training employed, the extra resources provided), the favorable findings capture only the impact of this program as a whole over the counterpart and cannot support the claim that students’ outstanding performance was caused by the inquiry-based approach used.

For the above reasons, we emphasize that there are differences between studies that examine inquiry as an instructional approach by separating out other contaminating factors and studies that vary levels of inquiry simultaneously with many other elements. Program-based studies are often conducted by educational researchers to compare the relative efficacy of an entire program or a curriculum over others. Such studies are important and that importance should be reflected in policy documents but the effectiveness of a program should not be attributed to any one of the designed specifics. It is not possible to derive conclusions concerning the efficacy of inquiry-based procedures (or for that matter any other procedure thus implemented) compared to more explicit instruction without at least some studies using a strict control of variables. Unfortunately, the main source of evidence in the field of science education about inquiry-based teaching comes from studies that more or less share the key characteristics discussed above. In a review of research on inquiry-based science instruction conducted between 1984 and 2002, Minner et al. (2010) found that out of the 138 studies from the final review list, 53% did not have a comparison group and

about another 25% used non-equivalent control groups. Their review indicated that many instructional components were included simultaneously when examining the impact of inquiry interventions and the levels of inquiry components included in the interventions varied substantially. Accordingly, the vast majority of studies examining inquiry-based science instruction are unable to serve as evidence supporting inquiry as an instructional approach.

Based on this work, it seems to us that the assumption that inquiry as an instructional approach improves students' learning has not been seriously tested since it was suggested decades ago as a theoretical idea. In 2002, Hofstein and Lunetta (2004) revisited their 1982 review (Hofstein & Lunetta, 1982) of research on inquiry-based laboratory science teaching. They stated in their 2002 review (published in 2004), "researchers have not comprehensively examined the effects of laboratory instruction on student learning and growth in contrast to other modes of instruction, and there is insufficient data to confirm or reject convincingly many of the statements that have been made about the importance and the effects of laboratory teaching" (p. 29). Twenty years later, in their 2002 review, these scholars stated, "...these comments are also valid at this writing 20 years later. That said, the assumption that laboratory experiences help students understand materials, phenomena, concepts, models, and relationships, almost independent of the nature of the laboratory experience, continues to be widespread in spite of sparse data from carefully designed and conducted studies" (p. 46). Today, almost another 20 years later, these researchers' concerns continue to echo. The use of inquiry learning as an instructional approach has been normalized on the assumption that the data overwhelmingly support that use. Many programs and curricula have been built using inquiry as an overarching instructional framework. Claims from field-initiated, program-based studies continue to be widely used as evidence to further promote inquiry approaches. An effective program does not indicate the effectiveness of any specific instructional approach, such as inquiry, that occurs within a program with many other elements. We need at least some studies demonstrating the advantages of inquiry-based learning over other approaches, using strict, randomized, controlled trials. Such controlled studies are widely available in the field of educational psychology but are frequently ignored when informing educational practice.

## Controlled Studies

Controlled studies use comparison groups and design intervention units that alter only one target factor at a time to examine the effectiveness of an instructional procedure over alternatives. This type of study commonly uses random assignment to ensure equivalent groups (within the bounds of probability) as opposed to matching. Because this type of study does not intend to examine the effectiveness of an entire program as a whole, it often does not design exemplar curricula or programs to simultaneously include all suggested instructional elements, such as technologies and teacher professional development, for testing. Rather, the intervention design focuses only on one target procedure at a time over alternatives. It needs to be emphasized that interactions between factors can be and are often tested using



factorial designs that simultaneously indicate the effect of two or more single factors and the interactions between them. The important point to note is that these studies allow assessment of the relative effectiveness of a single factor such as inquiry-based teaching as an instructional approach over other approaches.

With replications of experiments, it is possible to provide recommendations with considerable confidence (Mayer, 2003; Slavin, 2002). Whitehurst (2002, 2003), the then-director of Institute of Education Sciences (IES), argued that all evidence is not created equal and listed randomized, controlled trials as the “gold standard” for determining the effectiveness of instructional approaches. When faced with uncertainty concerning the effectiveness of competing approaches, Shavelson and Towne (2002) stated “[a] control group that has the same experiences as the experimental group except for the “treatment” under study is the best antidote” (p. 69). For these reasons, studies from the field of educational psychology heavily rely on this form of research to make instructional decisions.

Controlled studies have compared teaching involving exploration-based investigations with forms of direct explicit instruction, including simply giving students answers, having students watch demonstrations and listen to explanations, directly reading answers from texts, etc. These studies found that learners receiving these forms of direct, explicit instruction demonstrated greater learning of content and skills than those who learned through practicing investigations (Hsu et al., 2015; Klahr & Nigam, 2004; Matlen & Klahr, 2013; Renken & Nunez, 2010; Stull & Mayer, 2007; Zhang, 2018, 2019). A comprehensive analysis of 72 intervention studies examining students’ learning how to control variables during investigations (Schwchow et al., 2016) concluded that there is no evidence supporting the claim that teaching through inquiry-based, exploration-type investigations leads to better learning. Another comprehensive review of inquiry-based instruction by Lazonder and Harmsen (2016) concluded that students’ learning gain did not result from completing learning tasks by acting as scientists, but derived from direct explanations added to the program, which is clearly contradictory to the suggestion made in the Framework of the NGSS that “Core ideas...are specified not as explanations to be consumed by learners...” (National Research Council, 2012, p. 254). This trend is not surprising because there is a considerable number of findings using psychology of learning theories in favor of direct, explicit instruction when comparing it with exploration-based learning (Carlson et al., 1992; Kyun et al., 2013; Mayer, 2004; Moreno, 2004; Rittle-Johnson, 2006; Roussel et al., 2017; Tuovinen & Sweller, 1999). Evidence for the importance of direct, explicit instruction from a very large number of randomized, controlled studies suggests that for novice learners in a complex area, studying worked examples demonstrating the solutions of those problems has usually resulted in superior performance over conventionally used classroom practices finding the solution to the equivalent problems (Sweller et al., 2011, 2019).

There are various misunderstandings about the principles guiding controlled studies and their contributions to the field. The result is that a considerable number of such studies in educational psychology, like the ones listed above, have been left unattended in the formulation of science educational standards. Some argue that this type of study views instructional approaches dichotomously as either direct instruction or inquiry and is too simplistic to fully represent real classroom learning. We

reiterate that this type of study is not intended to provide exemplar curricula. Rather, they exclude contaminating elements to focus only on target procedures or factors over alternatives. The objection to scientifically rigorous experimental design is particularly ironic in the field of science education where controlled conditions and fair tests are extensively emphasized in the introduction of scientific methods to students. We argue that it is the very same reason and in the interests of scientific rigor that educational psychologists study learning by isolating factors and controlling conditions for comparisons.

Relatedly, some objections have been raised to studies that compare and prefer explicit instruction over inquiry-based science teaching by suggesting that these studies equate inquiry-based science teaching to pure discovery without acknowledging that in reality, an extensive amount of guidance including explicit instruction is provided to students. We emphasize that these studies do not create exemplar curricula interventions to include teachers' professional development, technologies, or argumentations that may sometimes include modules on explicit instruction, but only compare whether students learn better through practicing investigations or by being directly presented with the content. These studies simply suggest that having students practice investigations result in less understanding than having students directly interact with the content. We argue that such findings, mainly from controlled studies, at least warrant consideration in the formulation of educational policy. It must also be emphasized that just because an experiment only alters one factor at a time, it does not mean that the experimenter feels no other factor is important.

There is no more justification for excluding work based on randomized, controlled studies than there is for excluding work on exemplar curricula using program-based studies. As such, we must reiterate that our concern is not about the inclusion or use of program-based studies in policy documents, but solely on the complete exclusion of controlled studies from these documents.

Other objections are also raised to studies that alter one variable at a time on the grounds that classroom learning is complex with multiple variables interacting. This objection reflects a misunderstanding of the role of factorial experimental designs in determining the effects of interactions between factors. These studies do not violate the alter only one variable at a time rule. For example, the worked example effect occurs when learners presented worked examples to study learn more than those who learn by solving the equivalent problems (Chen et al., 2015, 2016). Nevertheless, based on the expertise reversal effect (Kalyuga et al., 2003), as levels of expertise in an area increase, the effect will decrease in size, then disappear, and eventually reverse with problem solving superior to studying worked examples. The effect can be demonstrated using a 2 (worked examples vs. problem solving)  $\times$  2 (lower expertise vs. higher expertise) experimental design with any statistically significant interaction effects explained by cognitive load theory (Paas et al., 2003).

Interactions between factors are important but they should not be solely handled by arbitrarily including a large number of random factors in an experiment. There are proper techniques for handling interactions and when used, the existence of interactions between factors does not violate the alter "one variable at a time rule." Statistical interactions were specifically designed to deal with complexity. There is no excuse for the blanket exclusion of all such studies.

Acknowledging the essential role of direct, explicit instruction in hands-on activities that promote investigations, research has examined the sequential ordering of investigation-based explorations and direct explicit instruction. A number of studies suggest an integration model where explorations begin the sequence followed by direct explicit instruction (DeCaro & Rittle-Johnson, 2012; Jacobson et al., 2017; Kapur, 2008, 2010; Kapur & Bielaczyc, 2012; Loibl & Rummel, 2014; Schwartz & Martin, 2004; Weaver et al., 2018). It has been argued that the exploration process motivates students and activates their pre-existing knowledge. When receiving subsequent instruction, students' prior problem-solving attempts enable them to better attend to critical concepts and allow them to see the subsequent instruction in a more meaningful and connected way (Kapur, 2008, 2010; Kapur & Bielaczyc, 2012). Interestingly, a recent meta-analysis by Loibl et al. (2017) reviewed various studies examining the approach of exploration followed by direct explicit instruction and found that this combination only works if the subsequent instruction specifically and directly explains the solutions to learners, which is parallel to what has been suggested in the direct, explicit instruction approach. In contrast, a number of studies support the relative effectiveness of direct explicit instruction first approach (Ashman et al., 2020; Fyfe et al., 2014; Glogger-Frey et al., 2015; Hsu et al., 2015; Matlen & Klahr, 2013; van Gog et al., 2011), arguing that starting with explorations overloads students' working memory (Kirschner et al., 2006; Sweller, 2009; Sweller et al., 2007).

Pending more research, Ashman et al., (2020) proposed a theoretical perspective to explain this inconsistency using the concept of task complexity. The complexity of a given task depends on learners' expertise. They argued that results favoring the exploration-first approach occur if learners are engaged in low complexity tasks. In contrast, when learners are involved in high complexity tasks, a direct explicit instruction first approach should be relatively effective.

Although there have been disagreements on the sequential orderings, none of these studies suggests that direct, explicit instruction is dispensable. Instead, these studies suggest direct explicit instruction is critical to science teaching and learning, a fact which does not seem to be acknowledged by standards over many years. In the absence of discussions addressing this issue in educational standards and reports, however, it is not uncommon to find teachers deliberately avoiding direct, explicit instruction in their attempts to implement instruction using investigations. For example, in observing lessons that adopt investigations, Furtak (2006) found that teachers often withhold concepts and procedures from students as a way to prioritize investigations. Abrahams and Millar (2008) also found that introducing students to concepts was often omitted in teaching that promoted investigations because teachers believe that the concepts and ideas will emerge from the act of investigation itself.

In summary, in the two previous sections, we presented contradictory results about teaching science through investigations by contrasting program-based studies with controlled studies. In the next section, we present a third methodology that has been used to address teaching through investigations.

## Correlational Studies

Recently, correlational studies also have joined the conversation and started to examine procedures that emphasized teaching science through investigations. Most of these studies draw from large international data sets and use statistical techniques to look for correlations between specific instructional elements and students' learning outcomes. Therefore, correlational studies differ from program-based studies that focus on testing the effectiveness of researcher-built inquiry-based programs as a whole. Moreover, those data sets tend to be a lot larger than those used by program-based studies. For example, TIMSS and PISA are two large initiatives collecting international data on students' performance and self-reported learning experiences, such as how often they design and conduct scientific investigations in schools. Thus, the findings from correlational studies can provide information about whether the dissemination of the policy-suggested investigation approach to science teaching is effective over large populations. Furthermore, using statistical models, correlational studies can disentangle students' learning experiences from one another. Coupled with multivariate designs to account for shared variance, correlational studies have unique access to associations between target learning experiences and students' performance within a large population. While correlational studies cannot function in the same manner as randomized, controlled trials, they have an advantage in that they can more easily use very large data sets and make relationships evident. Thus, this form of research can help identify factors and patterns. For these reasons, correlational studies are more objective than program-based studies.

In contrast to widespread positive findings from program-based studies, correlational studies using large data sets have found that the more students were involved in inquiry-based investigation activities, the more their science learning outcomes dropped. This pattern has emerged over years and across countries and regions at different levels of science performance, in the USA (Kaya & Rice, 2010; Zhang & Li, 2019), Finland (Lavonen & Laaksonen, 2009), Qatar (Areepattamannil, 2012), Australia, Canada, New Zealand (McConney et al., 2014), Chinese inner-Mongolia (Gao et al., 2017), England (Jerrim et al., 2019), Norway (Teig et al., 2018), Ireland, the UK, and the other 4 countries listed above (Oliver et al., 2019), 54 countries (Cairns & Areepattamannil, 2019), 10 best and 10 worst performing countries (Aditomo & Klieme, 2020), and 13 countries at similar science achievement levels (Forbes et al., 2020), and Taiwan (Liou, 2020). Importantly, student-led, inquiry-based investigation that has been mostly suggested by world-wide science teaching standards has been identified as the least effective instructional element. For example, using PISA 2015 results (OECD, 2016) on instructional factors associated with science performance, it was found that having students involved in various investigation steps, such as arguing about science questions and conducting investigations, had negative associations. Forbes et al. (2020) noted that "highly student-driven dimensions of inquiry, particularly procedural activities associated with investigation, are least frequently associated with high levels of student science achievement" (p. 19). Martin et al. (2020) tested the effects of cognitive load reduction instructional procedures on student engagement and achievement. Load reduction instruction includes explicit instruction for novice learners in an area. The authors surveyed

over 2000 secondary school science students concerning the extent to which load reduction instruction was used in their science classes and found a positive relation between load reduction instruction and engagement along with a positive relation between engagement and achievement. All these studies found that students' high achievement is associated with teacher-directed, explicit instructional strategies.

This clear pattern has been interpreted in different ways. Some have suggested implications for educational practice by arguing that the negative association is clear evidence that the field has not done enough to support inquiry teaching in schools and should place more resources for and emphasis on inquiry, while others have raised calls for re-evaluating inquiry as an instructional approach. For example, finding the negative correlations between inquiry activities and students' learning, Lavonen and Laaksonen (2009) insisted, "We do not suggest that the number of science inquiry activities should be reduced in Finland or in other OECD countries" (p. 937). Similarly, Aditomo and Klieme (2020) argued that "it would be misguided to use PISA findings to support arguments to scale back inquiry." In contrast, Zhang and Li (2019) argued that "there might be some grounds for questioning the current way of characterizing scientific investigations and call on researchers to further examine this topic" (p. 342).

It is important to note that such studies often use survey data to collect students' self-reported perceptions of what learning experiences they encounter, which are different from interventions that are strictly designed and consistently implemented across groups to test causal effects. As a result, correlational studies are limited in their capacity to claim causal relationships, which is also reflected in researchers' various interpretations. They can, however, suggest factors of interest which then can and need to be tested in subsequent controlled studies.

## Theoretical Issues Based on Cognitive Load Theory

In contrast to inquiry-based approaches, the explicit, structured approaches to instruction have strong, cognitively based theories underpinning them that are effective for generating predictions about learning gains and then explaining how and why these gains occurred. Cognitive load theory (Sweller et al., 2011, 2019) is one such theory and those aspects of the theory relevant to the current discussion will be briefly summarized. For categories of information relevant to instructional issues, the theory makes the following assumptions.

1. Humans have evolved to automatically obtain novel information either by problem solving or from other people. It is far more efficient to obtain information from others than to generate it oneself during problem-solving.
2. Once novel information is obtained, it must be processed by a limited capacity, limited duration working memory before being transferred to an unlimited long-term memory for later use. Compared to obtaining information from others, problem solving imposes a particularly heavy working memory load. Accordingly, we automatically gravitate to obtaining information from others. We engage in

problem solving when we do not have others from whom we can obtain novel information.

3. Once information is stored in long-term memory, it can be transferred back to working memory to govern action appropriate to a particular environment. Unlike when dealing with novel information, working memory has no known limits when dealing with familiar information transferred from long-term memory, resulting in the well-known transformative effects of education.

Based on cognitive load theory, there never is a justification for engaging in inquiry-based learning or any other pedagogically identical approaches when students need to acquire complex, novel information. As a species, we have evolved to obtain information from others and are particularly good at doing so. Data, especially from the worked example effect discussed above, provide strong empirical support, as do the correlation studies reviewed.

### Summary of the Three Research Paradigms

In brief, we have summarized three contrasting ways researchers have used to deal with the same issue—teaching science through inquiry—or exploration-based investigations. The three categories demonstrate that different forms of educational research appear to generate very different findings and conclusions that vary from supporting inquiry as an overarching approach in field-initiated, program-based studies, to consistent negative associations in correlation studies, and to overwhelming results questioning inquiry in controlled studies. These disparate findings are, however, not reflected proportionally in current standards and reports. It is obvious that selecting a different category of research will result in different implications for educational practice. When writers of educational standards choose to follow only one methodology and ignore the others, this choice could be at the expense of students' learning.

We argue that the success of future science education—and future scientists—relies on comprehensive standards that not only refer to a broader set of learning outcomes, but also represent a more balanced view of the available data. Thus, we further emphasize that none of the above should be seen as rejecting the use of any of the three research procedures discussed. But none of them alone should be used to definitively recommend any particular instructional procedure. Specifically, we should never use program-based studies as the sole source of evidence for any particular instructional procedure such as inquiry-based learning. All such recommendations should also include randomized controlled trials and large-scale correlational studies. However, program-based studies have been relied on almost exclusively in the standards to recommend inquiry-based learning with almost no questions addressed about the less favorable results from correlational and controlled studies. It is troubling to see sweeping curriculum reforms reinforced and overarching claims accepted while a large number of critical data sets have been ignored.

## Barriers and Reflection on Practice

To close this article, we would like to summarize some difficulties we have seen and hope our discussion can offer those who share our concerns an opportunity to reflect on current practices and to bridge the evidence gap between various forms of research, especially research conducted within an educational psychology framework.

Randomized, controlled studies are a reliable source when generating implications for educational practice (Campbell & Stanley, 1966; Hedges & Schauer, 2018; Levin & O'Donnell, 1999; Mayer, 2003; Shavelson & Towne, 2002; Slavin, 2002; Whitehurst, 2002, 2003) and are commonly used by educational psychologists. In a recent review, Hedges and Schauer (2018) noted that there has been a dramatic decline in the use of randomized, controlled trials in education in the USA between 1980 and 2000. Specifically, in the field of science education, Taylor et al. (2016) found that a vast majority of impact studies were not conducive to quantitative synthesis. A review of research about inquiry-based science instruction by Minner et al. (2010) found that about 78% of the included studies either did not have a comparison group or used non-equivalent control groups. Our review is consistent with these reviews, finding evidence that controlled studies are underrepresented in educational standards and reports.

We understand that to many educators, designing interventions that alter one factor only is not a fair representation of the inquiry approach. Instead, they advocate the use of program-based studies that design exemplar programs including all relevant factors. We reiterate the key differences between these two types of study as discussed above and argue that it is the very reason why controlled studies are essential and desperately needed in the field. In advocating for randomized, controlled studies, we emphasize that the use of such studies does not preclude the use of other research procedures but similarly; the use of those other research procedures should not preclude the use of randomized controlled trials and correlational studies when determining appropriate instructional designs.

Because the nature of randomized, controlled studies is to control for factors, some educators have concerns that this type of study is lab-based and cannot take the classroom context into account. In fact, it is entirely possible to conduct these studies in real classroom settings, as shown in many published studies discussed above, and that to provide the most reliable suggestions for educational practice, controlled studies need to be replicated across contexts by different research teams.

## Conclusions

Bridging the gap between research findings and educational practice can be difficult (Robinson et al., 2013; Wecker, 2013). We indicated that a large number of critical data sets have been neglected when formulating science educational policy, which has led to a disconnection between research findings and educational practice. We emphasize that none of the three research categories discussed should be rejected

or used exclusively when determining instructional procedures. All can contribute useful information. Unfortunately, program-based studies seem to be almost the only category used. While important, these studies should never provide the sole data sets when determining instructional procedures. There are dangers that this approach has led and will continue to lead to less than ideal educational results.

Although we have focused our discussion on students' understanding of science content and procedures, we acknowledge the importance of developing students' other learning outcomes, such as interests in and attitudes toward science. We reiterate that we have no objections to the various science learning outcomes. Nor do we reject any data indicating that investigation activities can be included to achieve these different learning goals. We wish to argue, however, that effectively developing students' understanding of science concepts and procedures should not be traded off by prioritizing other learning outcomes. It is hard to conceive of valid interests in and attitudes toward science without having the necessary conceptual knowledge and understanding.

There are many factors that are important to the success of science teaching. Elements, such as local contexts and resources, involvement of practitioners in collaborations, as well as nuances in the designed interventions, play important roles in the success of intervention implementations (Renkl, 2013, 2015; Wecker, 2013). However, we argue two points. First, randomly controlled trials have long been considered the only scientifically reliable procedure to establish a causal relationship. Second, we need to test whether such relationships can be effectively achieved in various contexts and what contextual elements and local practices affect the implementation. We believe the first point cannot be replaced by the second, or vice versa.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abrahams, I., & Millar, R. (2008). Does practical work really work? A study of the effectiveness of practical work as a teaching and learning method in school science. *International Journal of Science Education*, 30(14).
- Aditomo, A., & Klieme, E. (2020). Forms of inquiry-based science instruction and their relations with learning outcomes: Evidence from high and low-performing education systems. *International Journal of Science Education*. <https://doi.org/10.1080/09500693.2020.1716093>
- Areepattamannil, S. (2012). Effects of inquiry-based science instruction on science achievement and interest in science: Evidence from Qatar. *The Journal of Educational Research*, 105(2), 134–146.
- Ashman, G., Kalyuga, S., & Sweller, J. (2020). Problem-solving or explicit instruction: Which should go first when element interactivity is high? *Educational Psychology Review*, 32, 229–247. <https://doi.org/10.1007/s10648-019-09500-5>
- Australian Curriculum Assessment and Reporting Authority. (2018). Year 10 Australian Curriculum Retrieved from [www.australiancurriculum.edu.au/f-10-curriculum/science/aims/](http://www.australiancurriculum.edu.au/f-10-curriculum/science/aims/)



- Barton, A. C., & Tan, E. (2010). We be burnin'! agency, identity, and science learning. *Journal of the Learning Sciences, 19*(2), 187–229.
- British Columbia Ministry of Education. (2015). B.C.'s Curriculum: Science. Retrieved from <https://curriculum.gov.bc.ca/curriculum/science>
- Cairns, D., & Areepattamannil, S. (2019). Exploring the relations of inquiry-based teaching to science achievement and dispositions in 54 countries. *Research in Science Education, 49*(1), 1–23.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Houghton Mifflin Company.
- Carlson, R. A., Lundy, D. H., & Schneider, W. (1992). Strategy guidance and memory aiding in learning a problem-solving skill. *Human Factors, 34*(2), 129–145.
- Chen, O., Kalyuga, S., & Sweller, J. (2015). The worked example effect, the generation effect, and element interactivity. *Journal of Educational Psychology, 107*, 689–704.
- Chen, O., Kalyuga, S., & Sweller, J. (2016). Relations between the worked example and generation effects on immediate and delayed tests. *Learning and Instruction, 45*, 20–30.
- Coördinatiegroep curriculum.nu. (2019). Samen bouwen aan het primair en voortgezet onderwijs van morgen Deel 1: Adviezen van de Coördinatiegroep Curriculum.nu. [Building tomorrow's primary and secondary education together Part 1: Advice from the Curriculum.nu Coordination Group]. Den Haag, the Netherlands: Coördinatiegroep curriculum.nu. Retrieved from <https://www.curriculum.nu/download/advies-coordinatiegroep/>
- DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of Experimental Child Psychology, 113*(4), 552–568.
- Edelson, D. C., & Reiser, B. J. (2006). Making authentic practices accessible to learners: Design challenges and strategies. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 335–354). Cambridge University Press.
- Forbes, C. T., Neumann, K., & Schiepe-Tiska, A. (2020). Patterns of inquiry-based science instruction and student science achievement in PISA 2015. *International Journal of Science Education*. <https://doi.org/10.1080/09500693.2020.1730017>
- Furtak, E. M. (2006). The problem with answers: An exploration of guided scientific inquiry teaching. *Science Education, 90*(3), 453–467.
- Fyfe, E. R., DeCaro, M. S., & Rittle-Johnson, B. (2014). An alternative time for telling: When conceptual instruction prior to problem solving improves mathematical knowledge. *British Journal of Educational Psychology, 84*(3), 502–519.
- Gao, S., Wang, J., & Zhong, Z. (2017). Influence of science instruction reform on academic performance of eighth grade students in Chinese inner-Mongolia autonomous region. *Compare: A Journal of Comparative and International Education*. <https://doi.org/10.1080/03057925.2017.1365285>
- García-Carmona, A. (2020). From inquiry-based science education to the approach based on scientific practices. *Science & Education*. <https://doi.org/10.1007/s11191-020-00108-8>
- Geier, R., Blumenfeld, P. C., Marx, R. W., Krajcik, J. S., Fishman, B., Soloway, E., & Clay-Chambers, J. (2008). Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. *Journal of Research in Science Teaching, 45*(8), 922–939. Retrieved from - <https://doi.org/10.1002/tea.20248>
- Glogger-Frey, I., Fleischer, C., Grüny, L., Kappich, J., & Renkl, A. (2015). Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learning and Instruction, 39*(72–87).
- Hedges, L. V., & Schauer, J. (2018). Randomised trials in education in the USA. *Educational Research*. <https://doi.org/10.1080/00131881.2018.1493350>
- Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99–107.
- Hodson, D. (1996). Practical work in school science: Exploring some directions for change. *International Journal of Science Education, 18*(7), 755–760.
- Hofstein, A., & Lunetta, V. N. (1982). The role of the laboratory in science teaching: Neglected aspects of research. *Review of Educational Research, 52*(2), 201–217.
- Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education, 88*(1), 28–54. <https://doi.org/10.1002/sce.10106>
- Hsu, C.-Y., Kalyuga, S., & Sweller, J. (2015). When should guidance be presented during physics instruction? *Archives of Scientific Psychology, 3*(1), 37–53. <https://doi.org/10.3886/ICPSR35626.v1>
- Jacobson, M. J., Markauskaite, L., Portolese, A., Kapur, M., Lai, P. K., & Roberts, G. (2017). Designs for learning about climate change as a complex system. *Learning and Instruction, 1–14*.


- Jerrim, J., Oliver, M., & Sims, S. (2019). The relationship between inquiry-based teaching and students' achievement. New evidence from a longitudinal PISA study in England. *Learning and Instruction, 61*, 35–44.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23–31.
- Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26*, 379–424.
- Kapur, M. (2010). Productive failure in mathematical problem solving. *Instructional Science, 38*, 523–550.
- Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences, 21*(1), 45–83.
- Kaya, S., & Rice, D. C. (2010). Multilevel effects of student and classroom factors on elementary science achievement in five countries. *International Journal of Science Education, 32*(10), 1337–1363.
- Kirschner, P. A. (1992). Epistemology, practical work and Academic skills in science education. *Science & Education, 1*(3), 273–299.
- Kirschner, P. A. (2000). The inevitable duality of education: Cooperative higher education. Inaugural Address. Retrieved from Maarsticht, The Netherlands:
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86.
- Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction effects of direct instruction and discovery learning. *Psychological Science, 15*(10), 661–667.
- Kuhn, D. (2007). Is direct instruction an answer to the right question? *Educational Psychologist, 42*(2), 109–113.
- Kyun, S., Kalyuga, S., & Sweller, J. (2013). The effect of worked examples when learning to write essays in English literature. *The Journal of Experimental Education, 81*(3), 385–408.
- Lavonen, J., & Laaksonen, S. (2009). Context of teaching and learning school science in Finland: Reflections on PISA 2006 results. *Journal of Research in Science Teaching, 46*(8), 922–944.
- Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research, 86*(3), 681–718.
- Lee, H. S., & Anderson, J. R. (2013). Student learning: What has instruction got to do with it? *Annual Review of Psychology, 64*, 445–469.
- Levin, J. R., & O'Donnell, A. M. (1999). What to do about educational research's credibility gaps? *Issues in Education, 5*(2), 177–229.
- Liou, P.-Y. (2020). Students' attitudes toward science and science achievement: An analysis of the differential effects of science instructional practices. *Journal of Research in Science Teaching, Online First*, 1–25.
- Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review, 29*, 693–715.
- Loibl, K., & Rummel, N. (2014). The impact of guidance during problem-solving prior to instruction on students' inventions and learning outcomes. *Instructional Science, 42*, 305–326.
- Martin, A. J., Ginns, P., Burns, E. C., Kennett, R., & Pearson, J. (2020). Load reduction instruction in science and students' science engagement and science achievement. *Journal of Educational Psychology, Advance online publication*. <https://doi.org/10.1037/edu0000552>
- Matlen, B. J., & Klahr, D. (2013). Sequential effects of high and low instructional guidance on children's acquisition of experimentation skills: Is it all in the timing? *Instructional Science, 41*(3), 621–634.
- Mayer, R. E. (2003). Learning environments: The case for evidence-based practice and issue-driven research. *Educational Psychology Review, 15*(4), 359–366.
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*(1), 14–19. <https://doi.org/10.1037/0003-066x.59.1.14>
- McConney, A., Oliver, M. C., Woods-McConney, A., Schibeci, R., & Maor, D. (2014). Inquiry, engagement, and literacy in science: A retrospective, cross-national analysis using PISA 2006. *Science Education, 98*(6), 963–980.
- Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction—What is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching, 47*(4), 474–496.
- Mistler-Jackson, M., & Songer, N. (2000). Student motivation and Internet technology: Are students empowered to learn science? *Journal of Research in Science Teaching, 37*(5), 459–479.
- Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science, 32*, 99–113.
- National Academies of Sciences, Engineering, and Medicine. (2021). Call to action for science education: Building opportunity for the future. Washington, DC: The National Academies Press Retrieved from <https://doi.org/10.17226/26152>
- National Research Council. (1996). *National science education standards*. National Academy Press.

- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press.
- NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press.
- OECD. (2016). PISA 2015 results (Volume II). Policies and Practices for Successful Schools. Paris: OECD Publishing.
- Oliver, M., McConney, A., & Woods-McConney, A. (2019). The efficacy of inquiry-based instruction in science: A comparative analysis of six countries using PISA 2015. *Research in Science Education*. <https://doi.org/10.1007/s11165-019-09901-0>
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, 38(1), 1–4.
- Platform Onderwijs2032. (2016). Ons Onderwijs2032: Eindadvies [Our Education2032: Final advice]. The Hague, The Netherlands: Platform Onderwijs2032 Retrieved from <https://curriculum.nu/wp-content/uploads/2018/05/Ons-Onderwijs2032-Eindadvies-januari-2016.pdf>
- Renken, M. D., & Nunez, N. (2010). Evidence for improved conclusion accuracy after reading about rather than conducting a belief-inconsistent simple physics experiment. *Applied Cognitive Psychology*, 24(6), 792–811.
- Renkl, A. (2013). Why practice recommendations are important in use-inspired basic research and why too much caution is dysfunctional. *Educational Psychology Review*, 25(3), 317–324. <https://doi.org/10.1007/s10648-013-9236-0>
- Renkl, A. (2015). Different roads lead to Rome: the case of principle-based cognitive skills. *Learning: Research and Practice* 1(1), 79 - 90.
- Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, 77(1), 1–15.
- Robinson, D. H., Levin, J. R., Schraw, G., Patall, E. A., & Hunt, E. B. (2013). On going (way) beyond one's data: A proposal to restrict recommendations for practice in primary educational research journals. *Educational Psychology Review*, 25(2), 291–302.
- Roussel, S., Joulia, D., Tricot, A., & e., & Sweller, J. (2017). Learning subject content through a foreign language should not ignore human cognitive architecture: A cognitive load theory approach. *Learning and Instruction*, 52, 69–79.
- Schmidt, H. G., Loyens, S. M. M., Gog, T., & v., & Paas, F. (2007). Problem-based learning is compatible with human cognitive architecture: Commentary on Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42(2), 91–97.
- Schwartz, D. L., & Martint, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, 22(2), 129–184.
- Schwichow, M., Croker, S., Zimmerman, C., Höfler, T., & Härtig, H. (2016). Teaching the control-of-variables strategy: A meta-analysis. *Developmental Review*, 39, 37–63.
- Shaffer, D. W. (2004). Pedagogical praxis: The professions as models for postindustrial education. *Teachers College Record*, 106(7).
- Shavelson, R. J., & Towne, L. (2002). *Scientific Research in Education*. National Academy Press.
- Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, 31(7), 15–21.
- Songer, N. B., Lee, H.-S., & Kam, R. (2002). Technology-rich inquiry science in urban classrooms: What are the barriers to inquiry pedagogy? *Journal of Research in Science Teaching*, 39(2), 128–150.
- Songer, N. B., Lee, H.-S., & McDonald, S. (2003). Research towards an expanded understanding of inquiry science beyond one idealized standard. *Science Education*, 87(4), 490–516.
- Stull, A. T., & Mayer, R. E. (2007). Learning by doing versus learning by viewing: Three experimental comparisons of learner-generated versus author-provided graphic organizers. *Journal of Educational Psychology*, 99(4), 808–820.
- Sweller, J. (2009). What human cognitive architecture tells us about constructivism. In S. Tobias & T. M. Duffy (Eds.), *Constructivist instruction: Success or failure?* (pp. 127–143). Routledge.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.
- Sweller, J., Kirschner, P. A., & Clark, R. E. (2007). Why minimally guided teaching techniques do not work: A reply to commentaries. *Educational Psychologist*, 42(2), 115–121.
- Sweller, J., van Merriënboer, J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31, 261–292.
- Taylor, J., Furtak, E., Kowalski, S., Martinez, A., Slavin, R., Stuhlsatz, M., & Wilson, C. (2016). Emergent themes from recent research syntheses in science education and their implications for research design, replication, and reporting practices. *Journal of Research in Science Teaching*, 53(8), 1216–1231.

- Teig, N., Scherer, R., & Nilsen, T. (2018). More isn't always better: The curvilinear relationship between inquiry-based teaching and student achievement in science. *Learning and Instruction, 56*, 20–29.
- Tobias, S., & Duffy, T. M. (Eds.). (2009). *The success or failure of constructivist instruction*. Routledge.
- Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91*(2), 334–341.
- van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology, 36*(3), 212–218.
- Weaver, J. P., Chastain, R. J., DeCaro, D. A., & DeCaro, M. S. (2018). Reverse the routine: Problem solving before instruction improves conceptual knowledge in undergraduate physics. *Contemporary Educational Psychology, 52*, 36–47.
- Wecker, C. (2013). How to support prescriptive statements by empirical research: Some missing parts. *Educational Psychology Review, 25*(1), 1–18. <https://doi.org/10.1007/s10648-012-9208-9>
- Whitehurst, G. J. (2002). *Evidence-based education*. Paper presented at the Student Achievement and School Accountability Conference.
- Whitehurst, G. J. (2003). The Institute of Education Sciences: New wine, new bottles. Paper presented at the 2003 American Educational Research Association Annual Meeting Presidential Invited Session.
- Williams, M., & Linn, M. C. (2002). WISE inquiry in fifth grade biology. *Research in Science Education, 32*, 415–436.
- Zhang, L. (2016). Is inquiry-based science teaching worth the effort? Some thoughts worth considering. *Science & Education, 25*(7), 897–915.
- Zhang, L. (2018). Withholding answers during hands-on scientific investigations? Comparing effects on developing students' scientific knowledge, reasoning, and application. *International Journal of Science Education, 40*(4), 459–469. <https://doi.org/10.1080/09500693.2018.1429692>
- Zhang, L. (2019). "Hands-on" plus "inquiry"? Effects of withholding answers coupled with physical manipulations on students' learning of energy-related science concepts. *Learning and Instruction, 60*, 199–205. <https://doi.org/10.1016/j.learninstruc.2018.01.001>
- Zhang, L., & Li, Z. (2019). How does inquiry-based scientific investigation relate to the development of students' science knowledge, knowing, applying, and reasoning? An examination of TIMSS data. *Canadian Journal of Science, Mathematics and Technology Education, 19*(3), 334–345. <https://doi.org/10.1007/s42330-019-00055-9>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Lin Zhang<sup>1</sup>  · Paul A. Kirschner<sup>2,3</sup> · William W. Cobern<sup>4</sup> · John Sweller<sup>5</sup>

Paul A. Kirschner  
paul.kirschner@ou.nl

William W. Cobern  
bill.cobern@wmich.edu

John Sweller  
j.sweller@unsw.edu.au

<sup>1</sup> Department of Elementary & Special Education, 324 Harkins Hall, Providence College, 1 Cunningham Square, Providence, RI 02918, USA

<sup>2</sup> Open University of the Netherlands, Postbus 2960, 64901 DL Heerlen, The Netherlands

<sup>3</sup> Thomas More University of Applied Sciences, Sint-Andriesstraat 2, 2000 Antwerp, Belgium

<sup>4</sup> The Mallinson Institute for Science Education, Western Michigan University, Kalamazoo, MI, USA

<sup>5</sup> School of Education, University of New South Wales, Sydney, NSW 2052, Australia