

University of Groningen

Meta-analysis of inquiry-based learning

Lazonder, Ard W.; Harmsen, Ruth

Published in:
Review of Educational Research

DOI:
[10.3102/0034654315627366](https://doi.org/10.3102/0034654315627366)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2016

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research*, 86(3), 681-718. <https://doi.org/10.3102/0034654315627366>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Meta-Analysis of Inquiry-Based Learning: Effects of Guidance

Ard W. Lazonder and Ruth Harmsen
University of Twente

Research has consistently shown that inquiry-based learning can be more effective than other, more expository instructional approaches as long as students are supported adequately. But what type of guidance is adequate, and for whom? These questions are difficult to answer as most previous research has only focused on one type of guidance and one type of learner. This meta-analysis therefore synthesized the results of 72 studies to compare the effectiveness of different types of guidance for different age categories. Results showed facilitative overall effects of guidance on learning activities ($d = 0.66$, 95% CI [0.44, 0.88]), performance success ($d = 0.71$, 95% CI [0.52, 0.90]), and learning outcomes ($d = 0.50$, 95% CI [0.37, 0.62]). Type of guidance moderated the effects on performance success but not on the other two outcome measures. Considerable variation was found in the effects of guidance on learning activities, but the relatively low number of studies do not allow for any definitive conclusion on possible age-related differences.

KEYWORDS: inquiry learning, discovery learning, learner guidance, instructional support, age-appropriate support

Psychologists and educational scientists seem to converge on the notion that student involvement is key to successful learning (e.g., Freeman et al., 2014). The question of how students should be actively involved in the learning process is nevertheless still debated. In formal disciplines such as science and mathematics, a pedagogical approach known as inquiry learning qualifies as an organic way to make students active agents in their own learning process. Inquiry-based methods, in short, enable students to learn about a topic through self-directed investigations. By “acting like a scientist,” students not only learn science content but also science processes which both are included in the curriculum standards of many countries worldwide (Abd-El-Khalick et al., 2004; Achieve, 2010).

Despite their appealing nature, controversy remains as to whether and when inquiry-based methods promote student learning. Ausubel (1962), for example, contested that inquiry-based learning is only appropriate for learners in the concrete operational stage of development; once a person is capable of formal

operational reasoning it is more effective, and certainly more efficient, to directly teach the abstract principles underlying a topic or domain through expository methods. Taba (1963) on the other hand, emphasized the potential of inquiry learning to strengthen students' ability to use cognitive processes. Even though she admitted that it is unfeasible and undesirable to learn every topic through inquiry, she also believed that it would be possible to find the conditions of an effective teaching–learning strategy that would make inquiry learning possible.

These arguments have been repeated throughout the years, albeit in slightly different form, by scholars criticizing and advocating inquiry learning. In the most recent debate, Kirschner, Sweller, and Clark (2006) argued that inquiry-based methods are unlikely to be effective because they ignore the limitations of working memory (i.e., the cognitive structure in which conscious processing occurs). The mere conduct of an inquiry is already so demanding that it requires most cognitive capacity in working memory, leaving (too) little capacity to store novel information in long-term memory. A reply by Hmelo-Silver, Golan Duncan, and Chinn (2007) challenged this notion by arguing that contemporary inquiry-based methods are powerful and effective models of learning because they employ extensive scaffolding. Of interest is the shift in focus from inquiry learning per se to the conditions under which inquiry learning can be effective. As student guidance seems to be the key to answering this question, the present meta-analysis aims to contribute to this ongoing debate by investigating the conditions of effective inquiry learning support in science and math education.

The above discourse also points to the importance of defining what exactly should be understood by inquiry learning and when it can be considered “unguided,” “minimally guided,” or “guided.” As there appears to be no consistent definition of inquiry learning (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011; Klahr & Nigam, 2004; Strike, 1975), the present meta-analysis had to conceive one of its own, and defined the method as one in which students conduct experiments, make observations or collect information in order to infer the principles underlying a topic or domain. These investigations are governed by one or more research questions, either provided by the teacher or proposed by the student; adhere (loosely) to the stages outlined in the scientific method; and can be performed with computer simulations, virtual labs, tangible materials, or existing databases. This working definition is consistent with the description of inquiry learning used by the National Science Foundation (2000) and the scientific processes and practices outlined by the National Research Council (2012). The definition of guidance is more complicated and will be worked out in the sections that follow.

Previous Reviews of Inquiry Learning

Inquiry-based learning is rooted in the work of John Dewey (1859–1952), a philosopher of education who played a prominent role in educational reform in the first half of the 20th century. Dewey believed that instead of emphasizing the memorization of facts, science education should teach students how to think and act scientifically (National Research Council, 2000). Inquiry learning was adopted on a large scale in educational practice during the discovery learning movement in the 1960s, albeit for a different reason. Inspired by the work of Bruner (1961;

Bruner, Goodnow, & Austin, 1956), inquiry-based methods were initially embraced as a fruitful way to learn science content, but the influential work of scholars such as Kuhn (2005) and Klahr (2000) gradually renewed the emphasis on the cultivation of science process skills.

These echoes of time are reflected in the review studies on inquiry learning that have been published over the past decades. Early research syntheses by Bittinger (1968) and Hermann (1969) found inquiry-based learning to be more effective compared with expository forms of instruction. The advantage of inquiry learning was most apparent for the transfer of the learned material; superior retention was often obtained from expository methods. Bittinger's review further showed that these effects occur both in laboratory studies and in regular classrooms, although the effects in the latter setting tended to be less pronounced. In addition, Hermann found initial evidence that the effectiveness of inquiry learning depends on the guidance students receive during an inquiry. Despite considerable variation in the amount and type of guidance across studies, some within-study comparisons revealed that a reasonable degree of guidance is more effective than low guidance.

Thirty years later, T. De Jong and Van Joolingen (1998) reviewed the literature on simulation-based inquiry learning in science domains. Following an analysis of the typical problems students encounter in the various phases of the inquiry cycle, the authors synthesized empirical studies comparing the effectiveness of simulations with and without additional guidance (called "instructional measures"). Simulations enhanced with instructional measures were found to yield higher learning outcomes overall than simulations without this guidance. Three measures in particular appeared to be effective: giving learners access to domain information during the inquiry, offering assignments to structure the inquiry process, and constraining the complexity of the inquiry process through model progression. A review by T. De Jong, Linn, and Zacharia (2013) suggested that the advantages for guidance in simulation-based inquiry learning generalize to different instructional settings. Based on a synthesis of well-controlled comparison studies, the authors concluded that, for the acquisition of conceptual knowledge, investigations conducted with physical materials are as effective as technology-enhanced investigations that involve computer simulations or online laboratories.

A recent meta-analysis partially confirmed these findings (D'Angelo et al., 2014). Simulations were found to have an advantage in achievement over non-simulation instruction ($d = 0.62$), but did not affect students' learning activities ($d = 0.26$). These results should nevertheless be treated with some caution because the nonsimulation condition involved "some other kind of instructional treatment" (D'Angelo et al., 2014, p. 14), which does not rule out the possibility that the effect of using simulations was confounded with the instructional method. A more solid conclusion from this meta-analysis was that supplementing a simulation with some form of learner guidance had a modest effect on learning outcomes ($d = 0.49$) and inquiry skills ($d = 0.41$) but these effects proved independent of the type of guidance. However, two other variables did moderate the outcomes: fifth-graders failed to take advantage of the guidance offered via the simulation whereas older learners did, and guidance proved ineffective for simulation-based inquiry learning in math whereas significant positive effects were found in the domain of science.

Other recent meta-analyses have contrasted inquiry learning with more expository forms of instruction. The outcomes once again confirm the importance of learner guidance. Alfieri et al. (2011) found that across domains and settings, inquiry-based methods with minimal or no guidance are less effective than explicit instruction ($d = -0.38$). But when students receive adequate guidance during the inquiry, they learn more ($d = 0.30$) than students who are taught the same content by expository methods. Furtak, Seidel, Iverson, and Briggs (2012) conducted a similar meta-analysis in the domain of science. Students engaged in an inquiry with minimal guidance from their teacher tended to learn more than students exposed to “traditional instruction” ($d = 0.25$). However, studies comparing teacher-directed inquiry learning with traditional instruction had a considerably higher overall mean effect size ($d = 0.65$). The magnitude of this effect could not be replicated in the meta-analysis by Carolan, Hutchins, Wickens, and Cumming (2014), who found a small but significant overall benefit ($g = 0.15$) of more guided inquiry learning over less guided inquiry learning.

In conclusion, the research integrations presented in this section provide convincing evidence that inquiry-based methods can be more effective than other, more expository methods of instruction. Its effectiveness has mainly been demonstrated on learning outcomes assessed after the task by means of domain knowledge posttests; embedded assessment of the actions learners perform during an inquiry (i.e., learning activities) and the quality of the products they create during that inquiry (i.e., performance success) has received significantly less attention. The second conclusion that can be drawn is that the effectiveness of inquiry learning depends almost entirely on the availability of appropriate guidance. Which types of guidance are appropriate cannot be determined on the basis of the existing reviews and meta-analyses. This seems at least in part due to the fact that guidance is often classified ad hoc on the basis of the included studies. Using an a priori classification based on a theoretical framework might be more fruitful and ease interpretation of the findings. Finally, even though previous meta-analyses have paid attention to possible moderating effects due to the learners' age, the relative effectiveness of different types of guidance for different age groups has not yet been assessed.

These three conclusions shaped the design and focus of the present meta-analysis in which possible age-related differences in the effectiveness of inquiry learning guidance were examined. In order to contextualize the meta-analysis' research questions, the next section will give a more detailed account of what learners are supposed to do during an inquiry and propose various ways in which their activities can be supported.

Empirical Research on Inquiry Learning

Lazonder (2014) identified two major strands in inquiry learning research. On the one hand are studies investigating how particular groups of learners go about performing an inquiry. These studies often address developmental differences in scientific reasoning (i.e., the core thinking skills underlying successful inquiry learning), and offer the same, often minimal guidance across age groups to find out what learners of a certain age can and cannot do on their own account. The second line of research aims to identify the effects of different types of guidance

on learners' inquiry activities, performance success, and learning outcomes. These studies typically test the effects of one type of guidance with one homogeneous group of learners. The main findings of both lines of research are summarized below.

Developmental Differences in Scientific Reasoning

Scientific reasoning is defined as the application of scientific inquiry methods to reasoning situations (Kuhn & Franklin, 2006). A well-known and respected model describing the cognitive processes underlying scientific reasoning is the Scientific Discovery as Dual Search framework of Klahr and Dunbar (1988). According to this model, scientific discovery proceeds in iterative cycles of hypothesis generation, experimentation, and evidence evaluation. These three core processes have been extensively studied in the cognitive and developmental research conducted over the past 25 years.

The ability to *formulate hypotheses* is not present until elementary school age the earliest (Piekny & Maehler, 2013). From that age onward, children, teenagers, and adolescents all tend to begin their inquiry by focusing on hypotheses consistent with prior beliefs—but differ in their attempts to generate subsequent hypotheses. Cross-sectional research has shown that children tend to keep generating plausible hypotheses and often get stuck focusing on a single hypothesis, whereas older learners more often generate implausible hypotheses and are more likely to consider multiple hypotheses (Klahr, Fay, & Dunbar, 1993). Still, teenagers and even adolescents spontaneously generate very few hypotheses, and confuse hypotheses with predictions (Gijlers & De Jong, 2005; Njoo & De Jong, 1993). It thus seems that, although the ability to generate hypotheses has matured around the age of 12, inducing new or alternative hypotheses from data remains difficult across age groups (Klahr et al., 1993). This finding is particularly relevant to math education where learners mainly have to rely on available data or examples to create hypotheses (which are commonly referred to as conjectures). As mathematical objects and relations are abstract, they generally do not allow learners to make conjectures based on prior knowledge or everyday experience (Cañadas, Deulofeu, Figueiras, Reid, & Yevdokimov, 2007). Inducing conjectures from analogical cases or problems is perhaps an exception, but recognizing appropriate analogies can be challenging for learners (cf. Gick & Holyoak, 1980).

The ability to *design experiments* suitable for generating or testing hypotheses has proven to be difficult across ages. Five year olds are not yet able to distinguish between testing a hypothesis and generating an effect, but a notable increase in this ability occurs around the age of 6 (Piekny, Gruber, & Maehler, 2014; Sodian, Zaitchik, & Carey, 1991). Basic competence in carrying out unconfounded investigations of the relationship between variables that clearly covary can be acquired by many pupils by the age of 10 (Kanari & Millar, 2004; Schauble, Glaser, Duschl, Schulze, & John, 1995), but with the right support 6 and 7 year olds are already able to learn how to conduct sound investigations (Chen & Klahr, 1999; Varma, 2014). With age, children increasingly improve their ability to set up valid experimental comparisons and transfer this skill to new domains (Chen & Klahr, 1999; Koerber, Sodian, Kropf, Mayer, & Schwippert, 2011; Veenman, Wilhelm, & Beishuizen, 2004).

Evidence evaluation involves a critical assessment of the results obtained through experimentation, which should lead to informed decisions on whether and how hypotheses should be accepted, rejected, or revised and further examined. This coordination of theory and evidence often requires learners to make inferences about causal relationships between variables showing different patterns of covariation. The ability to evaluate perfect covariation and non-covariation evidence develops during the preschool and early elementary school years (Koerber et al., 2011; Piekny & Maehler, 2013). Preschool children at 4 years of age can already correctly interpret perfect covariation data as evidence supporting a causal hypothesis, and this ability improves significantly between the ages of 4 and 5 (Koerber, Sodian, Thoermer, & Nett, 2005; Piekny et al., 2014). The ability to interpret imperfect covariation seems to be comparatively more demanding (Inhelder & Piaget, 1958; Koslowski, 2008; Kuhn & Phelps, 1982). It is therefore no surprise that this skill does not seem to be well developed yet in early childhood (Koerber et al., 2005; Piekny et al., 2014), develops slowly, and hardly ever reaches maturity (Kuhn, Amsel, & O'Loughlin, 1988).

This summary of the research on scientific reasoning concurs with Zimmerman (2007), who concluded on the basis of a comprehensive review that “children are far more competent than first suspected, and likewise, adults are less so” (p. 213). Children from the age of 5 onward possess a basic ability to generate hypotheses, design and conduct experiments, and evaluate evidence, and therefore seem ready for inquiry learning. These young learners obviously need specific guidance in order for their inquiry to be successful, but the same could be said regarding older learners who, due to their higher levels of proficiency, can be assumed to require less specific types of support.

Learner Guidance to Support Scientific Reasoning

Studies investigating the effectiveness of inquiry learning in math and science education typically compare groups of learners who either do or do not receive a particular form of guidance. The results are generally in favor of guided inquiry learning, as was indicated by the meta-analyses presented above, but most studies provide insufficient evidence to explain these beneficial effects in light of the support needs of different age groups. Three important limitations are discussed below, together with their implications for the design of the present meta-analysis.

As knowledge acquisition is the ultimate goal of any instructional approach, inquiry learning research has predominantly and often exclusively focused on this particular outcome measure. This means that the effect of guidance is usually inferred from the knowledge learners have developed rather than assessed from the activities they perform during the inquiry. Although this type of research is defensive in its own right, a direct assessment of whether guidance promotes the inquiry processes as expected is more informative, and the studies that do combine process and outcome measures have indeed yielded valuable insight into the effects of the guidance under study (e.g., Burns & Vollmeyer, 2002; Chinn & Malhotra, 2002a; Lazonder, Hagemans, & De Jong, 2010; Toth, Klahr, & Chen, 2000).

Inquiry learning research also suffers from terminological inconsistency. Existing studies diverge considerably on which activities an inquiry comprises, and are equally inconsistent in designating the type of guidance being

investigated. Some researchers used colorful names such as *systematic science instruction* (Samarapungavan, Mantzicopoulos, & Patrick, 2008), *information tips* (Hulshof & De Jong, 2006), and *mechanistic cues* (Kaplan & Black, 2003) that do not convey the type of guidance learners receive. Other studies used different names for the same type of guidance. For example, the mechanistic cues used by Kaplan and Black (2003) served to remind learners to use the control of variables strategy (CVS) in designing experiments. In other writings, such reminders have been referred to as *prompts* (Marschner, Thillmann, Wirth, & Leutner, 2012), *heuristics* (Veermans, Van Joolingen, & De Jong, 2006), and *hints* (Rey, 2011). The reverse problem also occurs. Puntambekar and Hubscher (2005) noticed a change in the notion of scaffolding such that the term is increasingly being used to indicate support mechanisms that lack the core attributes of the original concept such as ongoing diagnosis, calibrated guidance, and fading. This lack of operational definitions complicates the comparison of the efficacy of different types of guidance and is likely to have caused unwarranted disagreement among scholars in the field (Klahr, 2013).

Terminological differences aside, previous research has almost exclusively focused on one type of guidance and one type of learner. Remarkably, however, the researchers' choice of specific or less specific guidance appears to be independent of the learners' age. Explicit support such as direct instruction of the CVS has been successfully applied with children (e.g., Klahr & Nigam, 2004; Varma, 2014), but older learners benefit from explicit directions too (e.g., Sao Pedro, Gobert, Heffernan, & Beck, 2009). Similarly, less specific types of support such as model progression have been found to promote inquiry learning in children (White, 1993), teenagers (Wichmann & Timpe, 2013), and adolescents (Mulder, Lazonder, & De Jong, 2011). As few studies have investigated the effects of one type of guidance across age groups, the present meta-analysis aimed to synthesize the results of individual studies to compare the effectiveness of different types of support for different age categories.

The limitations outlined above point to two implications for the design of this meta-analysis. One is that the scope of the analysis should go beyond the mere assessment of learning outcomes. Although it may be difficult to include enough studies investigating how guidance shaped the inquiry learning process, this information is a valuable addition to the empirical research and makes the meta-analysis stand apart from previous research integrations. The second implication is that the meta-analysis should start from a typology of inquiry learning guidance that differentiates different types of guidance based on their specificity.

Typology of Inquiry Learning Guidance

As a first step in investigating whether a differential effect of inquiry learning guidance indeed exists, this meta-analysis defined *guidance* as any form of assistance offered before and/or during the inquiry learning process that aims to simplify, provide a view on, elicit, supplant, or prescribe the scientific reasoning skills involved. As both the term guidance and its definition are rather broad, some further specification seems appropriate. In previous writings, guidance has often been classified in terms of the learning activities it aims to support. T. De Jong (2006), for example, organized various types of guidance according to the

phases of the inquiry cycle, and Lazonder (2014) did the same for the three core scientific reasoning processes. A more coarse-grained classification was proposed by Reid, Zhang, and Chen (2003), who made a distinction between interpretative support that helps learners understand important domain concepts, experimentation support for guiding learners in designing and conducting experiments, and reflective support that assists learners in looking back on their inquiry and the knowledge acquired. A similar classification was proposed by Quintana et al. (2004) who presented a series of scaffolding principles and guidelines to support sense making, process management, and articulation and reflection.

Even though the above classifications are generally acknowledged and frequently cited in the research literature, they are less appropriate for the purpose of this meta-analysis. Developmental research suggests that younger, less experienced learners need more explicit guidance than older learners; investigating this conjecture thus requires a framework that classifies guidance in terms of extensiveness or explicitness rather than the skills or activities involved. A potentially more appropriate typology was proposed by T. De Jong and Lazonder (2014) who organized their framework according to the specificity of the guidance learners need to successfully perform an inquiry. This typology was used in the present meta-analysis; a short description of its defining characteristics is given in Table 1.

Process constraints offer the least specific type of guidance and are therefore intended for learners with matured inquiry skills. Process constraints give no overt directions at all but instead organize the inquiry into a series of manageable subtasks, for instance, by increasing the number of elements the learner should investigate (White, 1993), increasing the fidelity of the learning environment (Alessi, 1995), or increasing the number of features the learner can control (Rieber & Parmley, 1995). *Status overviews* are more specific in that they summarize what or how well the learner has performed; the decision to use this data to maintain or adapt one's behavior is entirely up to the learner. An example of a status overview in collaborative inquiry learning is the Participation Tool that visualizes how much group members have contributed to the online learning dialogue. The use of this tool was found to encourage the more silent or less active students to increase their participation (Janssen, Erkens, Kanselaar, & Jaspers, 2007). *Prompts* are timed cues, either given by a human being or embedded within the learning environment, that remind the learner to perform a particular action. Prompts are more specific than status overviews because they tell the learner what to do (but not how to do it) at appropriate moments during the inquiry.

The remaining types of guidance all provide the learner with guidelines on how to perform a certain activity, but differ with regard to the specificity of the directions given. *Heuristics* remind learners to perform an action *and* point out possible ways to perform that action. Similar to prompts, heuristics can be implemented as cues in the learning environment that appear either at preset points in the inquiry (e.g., Lee & Chen, 2009) or in adaptive response to the learner's actions (e.g., Veermans et al., 2006). Alternatively, the full set of heuristics can be given at the outset of the inquiry, for example as a series of assignments or as process guidelines on student worksheets (e.g., De Vries, Van der Meij, & Lazonder, 2008; Njoo & De Jong, 1993).

TABLE 1*Typology of inquiry learning guidance*

Type of support	Basic idea	Intended audience
Process constraints	Restrict the comprehensiveness of the learning task	Learners who are able to perform and regulate the basic inquiry process, but still lack the experience to do so under more demanding circumstances
Status overviews	Make task progress or learning visible	Learners who are able to perform the basic inquiry process, but lack the skills to plan and keep track of their learning trajectory
Prompts	Remind to perform an action	Learners who are able to perform an action but may not do so on their own initiative
Heuristics	Remind to perform an action and suggest how to perform that action	Learners who do not know exactly when and how an action should be performed
Scaffolds	Explain or take over the more demanding parts of an action	Learners who do not have the proficiency to perform an action themselves or cannot perform the action from memory
Explanations	Specify exactly how to perform an action	Learners who are (largely) incognizant of the action and how it should be performed

Note. Based on T. De Jong and Lazonder (2014), with minor textual modifications.

Scaffolds offer more specific guidance than heuristics: they assist learners in performing demanding activities by explaining what to do and how to do it, and provide designated means to carry out, structure, or simplify the learner's actions. Consistent with Vygotsky's (1978) notion of the zone of proximal development, scaffolds explain or take over the difficult parts of the activity; when the learner's skill level increases, the scaffolding is gradually removed so that the learner eventually performs the activity without assistance. However, this fading is often and rightfully absent in short-term inquiries because inquiry skills develop slowly. Scaffolds come in many forms and have been designed for different age groups. For example, the Go-Lab environment (T. De Jong, Soteriou, & Gillet, 2014) incorporates scaffolds that assist learners in elementary education, middle school and high school during the key phases of an inquiry such as formulating research questions, generating hypotheses, and designing experiments. For each of these activities a designated scaffold is available (see <http://www.golabz.eu/apps>).

Explanations offer the most specific type of guidance and are therefore intended for learners who lack the basic ability to perform an inquiry skill. Unlike all other types of guidance, explanations can be given either before the inquiry, for

instance as a short preparatory training, or during the inquiry on a just-in-time basis. One of the most well-known examples concerns the direct instruction in experimental design developed by Klahr and co-workers (see, for an overview, Klahr & Li, 2005). Their CVS instruction proved highly successful in promoting children's ability to design unconfounded experiments, and yielded sustained effects until 3 years after the instruction (Strand-Cary & Klahr, 2008). In all these studies, children learned the basics of experimental design through short teacher-led instructions, examples, and guided practice. More recently, the principles underlying this CVS instruction have been used to design a software tutor where explanations on the use of the CVS are embedded within the learner's inquiry (e.g., Siler, Klahr, Strand-Cary, & Magaro, 2009).

Research Questions and Hypotheses

This meta-analysis aimed to answer three questions: (a) what is the effectiveness of inquiry learning guidance on learning activities, performance success, and learning outcomes, respectively; (b) does this effectiveness depend on the type of guidance; and (c) does the effectiveness of different types of guidance depend on the learner's age? As previous meta-analyses by Carolan et al. (2014) and D'Angelo et al. (2014) reported positive effects of guided inquiry learning over unguided inquiry learning, the present meta-analysis was expected to yield similar outcomes. Consistent with the view of Kirschner et al. (2006), this overall positive effect was expected to differ as a function of the type of guidance, with larger effect sizes being associated with more specific types of guidance. However, the typology of guidance described in the preceding section suggests that specific forms of guidance such as scaffolds or explanations might be comparatively more effective for younger learners who experience more difficulties during an inquiry than older learners, who in turn might benefit more from less specific guidance.

Method

Literature Search

Studies investigating guidance in inquiry learning were identified through a search of reference databases and a perusal of relevant conference proceedings. First, the databases of ERIC, PsycINFO, and Web of Science were consulted for the period 1993 to 2013. This time span was selected in order to include exactly two decades of recent research on inquiry-based learning. After several trial runs, the final searches were conducted in January 2014 with the following queries: [("inquiry learning" OR "discovery learning" OR "control of variables") AND ("students" OR "support" OR "instruction")], ["scaffold*" AND "inquiry"], [{"scientific thinking" OR "scientific reasoning"} AND ("learning" OR "instruction")]. The Web of Science search, limited by SSCI and CPI-SSH citation indexes, resulted in 738 hits. The searches in ERIC and PsycINFO, additionally limited by [AB abstract], returned 774 and 589 hits, respectively. By comparing the total of 2,101 hits, 524 duplicates were found; exclusion of those duplicates resulted in 1,577 unique hits.

Second, online conference proceedings of EARLI (European Association for Research on Learning and Instruction), NARST (National Association of Research in Science Teaching), and AERA (American Educational Research Association)

were searched for papers concerning inquiry learning guidance. This yielded an additional 89 hits which were collected online (29) or requested from the authors (63 requested, 31 responses, 27 papers received). The 56 retrieved papers were scanned for duplicates with the search results from the databases, which resulted in the removal of 4 papers and brought the total number of available and unique conference papers to 52.

Inclusion of Studies

To be eligible for inclusion in this meta-analysis, a study had to (a) investigate inquiry-based learning in math, physics, chemistry, biology, or general science with learners aged 5 to 22; (b) compare a group of learners that received a particular type of guidance against a reference group that did not receive this guidance—regardless of whether additional basic learning support was available to learners in both groups; (c) either randomly assign learners to one of these two groups, confirm the comparability of both groups in preliminary analyses, or control for possible preexisting differences in the main analyses; and (d) assess the effects of the guidance under study on participants' learning activities, performance success, or learning outcomes, and report these effects quantitatively by means of descriptive or inferential statistics.

The first three criteria were used in an initial screening of the studies' abstracts. If no abstract was available, the full publication was collected and examined. This first round of selection resulted in the provisional inclusion of 216 studies. In order to reach a final decision, these studies were retrieved from an online library or requested from the authors. The 213 studies that were eventually obtained were read by both authors for inclusion. For studies with insufficient statistical information to compute an effect size, the main author was contacted with a request to provide the missing data. If the author did not respond or could not provide the requested information, the study was discarded. After differences in judgment were discussed, the authors agreed that 68 studies met all four inclusion criteria. As 4 of these investigations reported on two experiments with separate samples, the final number of studies included in this meta-analysis was 72 (see Table 2).

Outcome Measures and Possible Moderators

Outcome measures included in this meta-analysis were learning activities, performance success, and learning outcomes. *Learning activities* pertained to what participants did during the inquiry. These actions included a variety of skills ranging from planning the inquiry process through generating hypotheses and designing experiments to reflecting on outcomes, which were assessed from the experimenter's notes, the learners' oral or written statements, or computer log files. *Performance success* indicated what learners managed to achieve during the inquiry, as evidenced by the products they created throughout the learning process. Examples include the number of valid inferences, the quality of a concept map, and the proportion of correctly completed assignments. *Learning outcomes* were assessments administered immediately or shortly after the learning process. These measurements sought to determine what participants had learned from the inquiry through posttests, criterion tasks, interviews, or questionnaires.

TABLE 2*Studies included in this meta-analysis*

Author(s)	Publication type	Type of guidance	Age group	Domain	Outcome measure and focus ^a	Study design ^b
Abdullah and Shariff (2008)	Journal (IF < 1.5)	Scaffolds	Adolescents	Science	LO-is	RE
Baroodly, Eiland, Purpura, and Reid (2012)	Journal (IF < 1.5)	Scaffolds	Children	Math	LO-dk	RE
Baroodly, Eiland, Purpura, and Reid (2013)	Journal (IF ≥ 1.5)	Scaffolds	Children	Math	LO-dk	RE
Chen and Klahr (1999)	Journal (IF ≥ 1.5)	Explanations	Children	Science	LO-is	RE
Chin, Dohmen, and Schwartz (2013)	Journal (IF < 1.5)	Scaffolds	Children	Science	LO-dk	QE
Davis (2000, Exp. 1)	Journal (IF < 1.5)	Prompts	Teenagers	Science	PS-dk	QE
Davis (2000, Exp. 2)	Journal (IF < 1.5)	Prompts	Teenagers	Science	PS-dk	QE
K. De Jong and Brand-Gruwel (2008)	Journal (IF < 1.5)	Scaffolds	Children	Science	PS-dk, LO-dk	QE
T. De Jong et al. (1999)	Journal (IF ≥ 1.5)	Heuristics	Adolescents	Science	LO-dk	QE
Dean and Kuhn (2007)	Journal (IF ≥ 1.5)	Explanations	Children	Science	LO-is	QE
Dejonckheere, Van de Keere, and Mestdagh (2009, Exp.2)	Journal (IF < 1.5)	Explanations	Children	Science	LO-dk	QE
Dejonckheere, Van de Keere, and Tallir (2011, Exp.2)	Journal (IF < 1.5)	Explanations	Children	Science	LO-is	QE
Dyke, Adamson, Howley, and Rose (2013)	Journal (IF < 1.5)	Scaffolds	Teenagers	Science	LO-dk	RE
Eckhardt, Urhahne, Conrad, and Harms (2013)	Journal (IF ≥ 1.5)	Scaffolds	Teenagers	Science	LO-dk	QE
Fund (2007)	Journal (IF ≥ 1.5)	Prompts	Teenagers	Science	LO-dk	QE
Gijlers and De Jong (2009)	Journal (IF < 1.5)	Scaffolds	Adolescents	Science	LA-is, LO-dk	RE
Gijlers and De Jong (2013)	Journal (IF ≥ 1.5)	Scaffolds	Adolescents	Science	LA-is, LO-dk	RE
Graesser et al. (2007, Exp. 1)	Journal (IF ≥ 1.5)	Scaffolds	Adolescents	Science	LA-is, LO-dk	RE

(continued)

TABLE 2 (continued)

Author(s)	Publication type	Type of guidance	Age group	Domain	Outcome measure and focus ^a	Study design ^b
Hagemans, Van der Meij, and De Jong (2013, Exp. 1)	Journal (IF ≥ 1.5)	Status overviews	Adolescents	Science	LA-rs, PS-dk, LO-dk	QE
Hagemans et al. (2013, Exp. 2)	Journal (IF ≥ 1.5)	Status overviews	Teenagers	Science	LA-rs, PS-dk, LO-dk	QE
Han and Black (2011)	Journal (IF ≥ 1.5)	Explanations	Children	Science	LO-dk	QE
Hogan (1999)	Journal (IF ≥ 1.5)	Explanations	Teenagers	Science	LO-is	QE
Huang (2011, Exp. 1)	Dissertation	Scaffolds	Teenagers	Science	LO-dk	QE
Huang (2011, Exp. 2)	Dissertation	Scaffolds	Teenagers	Science	LO-dk	QE
Hulshof and De Jong (2006)	Journal (IF < 1.5)	Explanations	Adolescents	Science	LA-is, LO-dk	RE
Hushman (2011)	Dissertation	Explanations	Children	Science	LA-is, LO-is	RE
Keselman (2003)	Journal (IF ≥ 1.5)	Explanations	Adolescents	Science	LO-is	QE
Knaggs and Schneider (2012)	Journal (IF < 1.5)	Scaffolds	Children	Science	PS-dk	RE
Kolloffel, Eysink, and De Jong (2010)	Journal (IF ≥ 1.5)	Scaffolds	Adolescents	Math	LO-dk	RE
Kramarski and Dudai (2009)	Journal (IF < 1.5)	Explanations	Teenagers	Math	LO-is	QE
Kuhn and Dean (2005)	Journal (IF ≥ 1.5)	Prompts	Children	Science	LO-is	RE
Kuhn and Katz (2009)	Journal (IF ≥ 1.5)	Prompts	Children	Science	LO-is	QE
Kunsting, Kempf, and Wirth (2013)	Journal (IF ≥ 1.5)	Explanations	Teenagers	Science	LA-is, LO-dk	RE
Lazonder and Egberink (2014) ^c	Journal (IF ≥ 1.5)	Process constraints	Children	Science	LA-is, PS-dk, LO-is	QE
Lazonder and Kamp (2012)	Journal (IF ≥ 1.5)	Process constraints	Children	Science	LA-is, PS-dk, LO-dk	RE
Lederman, Lederman, Wickman, and Lager-Nyqvist (2007)	Conference paper	Explanations	Teenagers	Science	LO-dk	QE
Lin and Lehman (1999)	Journal (IF ≥ 1.5)	Prompts	Adolescents	Science	LO-is	RE

(continued)

TABLE 2 (continued)

Author(s)	Publication type	Type of guidance	Age group	Domain	Outcome measure and focus ^a	Study design ^b
Lorch et al. (2010)	Journal (IF ≥ 1.5)	Explanations	Children	Science	LA-is, LO-is	QE
Lorch et al. (2014) ^c	Journal (IF ≥ 1.5)	Explanations	Children	Science	LA-is, LO-is	QE
MacGregor and Lou (2005)	Journal (IF < 1.5)	Scaffolds	Children	Science	PS-dk, LO-dk	QE
Mandrin and Preckel (2009)	Journal (IF < 1.5)	Scaffolds	Teenagers	Math	LO-dk	RE
Manlove, Lazonder, and De Jong (2006)	Journal (IF ≥ 1.5)	Scaffolds	Adolescents	Science	LA-rs, PS-dk	RE
Manlove, Lazonder, and De Jong (2007)	Journal (IF ≥ 1.5)	Scaffolds	Adolescents	Science	PS-dk	RE
Mannheimer Zydney (2005)	Journal (IF ≥ 1.5)	Scaffolds	Adolescents	Science	LA-is, PS-dk, LO-dk	QE
Marschner et al. (2012)	Journal (IF < 1.5)	Heuristics	Teenagers	Science	LA-is	RE
Matlen and Klahr (2013)	Journal (IF ≥ 1.5)	Explanations	Children	Physics	LO-is	RE
Milne et al. (2010)	Conference paper	Explanations	Adolescents	Science	LO-dk	RE
Milne et al. (2011)	Conference paper	Explanations	Adolescents	Science	LO-dk	QE
Minogue and Borland (2012)	Conference paper	Explanations	Adolescents	Science	LO-dk	RE
Moreno (2004, Exp. 1)	Journal (IF ≥ 1.5)	Explanations	Adolescents	Science	LO-dk	RE
Moreno (2004, Exp. 2)	Journal (IF ≥ 1.5)	Explanations	Adolescents	Science	LO-dk	RE
Morgan and Brooks (2012)	Journal (IF < 1.5)	Prompts	Adolescents	Science	LO-is	RE
Mulder et al. (2011)	Journal (IF ≥ 1.5)	Process constraints	Adolescents	Science	PS-dk	RE
Mulder, Lazonder, De Jong, Anjewierden, and Bollen (2012)	Journal (IF < 1.5)	Process constraints	Adolescents	Science	LA-is, PS-dk	RE
Neber and Anton (2008)	Journal (IF < 1.5)	Explanations	Teenagers	Science	LO-dk	QE
Plass et al. (2013, Exp. 2)	Conference paper	Scaffolds	Children	Science	LO-dk	RE
Raes, Schellens, De Wever, and Vanderhoven (2012)	Journal (IF ≥ 1.5)	Heuristics	Adolescents	Science	LO-dk	QE

(continued)

TABLE 2 (continued)

Author(s)	Publication type	Type of guidance	Age group	Domain	Outcome measure and focus ^a	Study design ^b
Reid, Zhang, and Chen (2003)	Journal (IF ≥ 1.5)	Scaffolds	Teenagers	Science	LO-dk	RE
Roll, Holmes, Day, and Bonn (2012)	Journal (IF ≥ 1.5)	Prompts	Adolescents	Science	LA-rs, PS-dk	QE
Saab and Van Joolingen (2005)	Conference paper	Scaffolds	Adolescents	Science	LO-dk	RE
Sao Pedro, Gobert, and Sebuwufu (2011)	Conference paper	Explanations	Teenagers	Science	LO-is	RE
Steffensky, Langkes, Carstensen, and Nölke (2012)	Journal (IF < 1.5)	Explanations	Children	Science	LO-dk	QE
Swaak, Van Joolingen, and De Jong (1998)	Journal (IF ≥ 1.5)	Process constraints	Adolescents	Science	LA-is, LO-dk	RE
Van der Meij, Van der Meij, and Harmsen (2013)	Conference paper	Prompts	Teenagers	Science	LO-dk	RE
Veenman, Elshout, and Busato (1994)	Journal (IF ≥ 1.5)	Prompts	Adolescents	Science	LO-dk	RE
Vreman-de Olde and De Jong (2006)	Journal (IF ≥ 1.5)	Heuristics	Adolescents	Science	PS-dk	RE
Wichmann and Leutner (2009)	Journal (IF < 1.5)	Prompts	Adolescents	Science	LA-is, LO-dk	RE
Wichmann and Timpe (2013)	Conference paper	Process constraints	Teenagers	Science	LO-dk	RE
Wollenschlager, Moller, and Harms (2011)	Journal (IF < 1.5)	Explanations	Teenagers	Science	PS-dk	RE
Zhang (2013)	Journal (IF ≥ 1.5)	Scaffolds	Teenagers	Science	LA-is	QE
Zohar and Ben David (2008)	Journal (IF ≥ 1.5)	Explanations	Teenagers	Science	LO-is	RE
Zohar and Peled (2008)	Journal (IF ≥ 1.5)	Explanations	Children	Science	LO-is	RE

Note. Some studies incorporate multiple types of guidance or multiple outcome measures; the ones listed here were included in the meta-analysis for reasons outlined in the Method section. IF = journal impact factor.

^aLA = Learning Activities; PS = Performance Success; LO = Learning Outcomes; is = inquiry skills; rs = regulative skills; dk = domain knowledge. ^bRE = randomized experiment; QE = quasi-experiment (with nonequivalent control group design). ^cRetrieved as advance online publication; published in 2014.

To explore possible differential effects on these outcome measures, seven moderators were extracted from the studies. The first moderator, *outcome focus*, specified in more detail what each outcome measure addressed. This could be inquiry skills, regulative skills, or domain knowledge, depending on the outcome measure in question. Learning activities for example, was defined as a measurement of skill and could therefore only pertain to inquiry skills and regulative skills. Likewise, measures of performance success could only target domain knowledge whereas learning outcomes could refer to all three types of outcome focus.

The second moderator, *publication type*, referred to whether a study appeared as a journal article, conference paper, or dissertation. Journal articles were further classified according to the 2013 release of Thompson Reuters' Journal Citation Reports®. Studies from journals with an impact factor of 1.5 or higher were coded as first-tier journal article; all other journal publications were considered second-tier articles. If studies from conference papers and dissertations had been reworked into a journal article, they were coded as such if the publication date fell within the time span covered by this meta-analysis.

The third moderator was the *domain* of the study, which concerned the disciplinary nature of the topic of inquiry. A coarse-grained distinction was made between math and science, where math was used as a general designation for disciplines such as arithmetic, geometry, algebra, and statistics, and science served as overarching term for physics, chemistry, and biology. Fourth, the *study's design* was classified as either randomized experiment or quasi-experiment. The former type of design is characterized by a random allocation of students across experimental conditions whereas the latter lacks such randomization, often because practical constraints require researchers to work with intact classes. *Duration* was the fifth moderator. It provided a rough indication of the time learners' spent on their inquiry, and hence did not include the time for pre- or post-session assessments. A dichotomous classification was used to differentiate studies that included a single session from studies with multiple sessions. Single-session studies typically lasted 20 to 100 min; multiple-session studies comprised two or more of these lessons that took place on consecutive days or with a delay of 1 week maximum.

The final two moderators represent the key factors in this meta-analysis. *Type of guidance* indicated the specificity of the support given to learners in the experimental condition. Consistent with the typology presented in Table 1, a distinction was made between process constraints, status overviews, prompts, heuristics, scaffolds, and explanations. These category labels reflect the terminology used in this meta-analysis and do not necessarily match the designation used in the original study. With *age group*, the study's sample was classified as children, teenagers, or adolescents. Children were 5 to 12 years of age, teenagers were 12 to 15 years of age, and adolescents were 15 to 22 years of age. This coding occurred on the basis of the sample's mean age. In case the sample's mean age or age range was not provided, it was inferred from participants' grade level, taking into account the differences in educational systems across countries.

Internal consistency of the outcome measures extraction and coding was determined by having both authors score all studies independently. Interrater

agreement on learning activities, performance success, and learning outcomes was high with an overall Cohen's κ of .91. All disagreements were resolved by discussion. Concerning moderator coding, straightforward study features such as publication type, domain, study design, and duration were scored by the second author, who consulted the first author when in doubt. Coding of study features that required a subjective interpretation of the rater were preceded by interrater reliability assessments. Toward this end, both authors independently scored a set of 33 randomly selected studies. The Cohen's κ agreement estimates were .83 (type of guidance), .91 (age group), and .91 (outcome focus). After disagreements were resolved by discussion, the second author coded the remaining studies.

Computation of Effect Sizes

Standardized mean differences were computed for each outcome measure. For learning activities and performance success, this effect size metric was defined as the difference between the mean scores of the treatment group and the control group divided by the pooled standard deviation. When means or standard deviations were not reported, effect sizes were obtained from inferential statistics (t , χ^2 , and F) or raw data. A similar procedure was used for learning outcomes except that, where possible, effect sizes were based on gain scores so as to compensate for potential a priori differences between conditions. In studies that provided both pretest and posttest data, the mean gains and standard deviations were calculated via the conversion formula provided by Lipsey and Wilson (2001). If neither gains nor pre-post scores were available, the means and standard deviations of the posttest scores were used to calculate the effect size. Effect size estimates of all outcome measures were corrected for small sample bias using the procedure by Hedges and Olkin (1985). This unbiased effect size index will be indicated by the parameter d throughout this meta-analysis.

The methods proposed by Borenstein, Hedges, Higgins, and Rothstein (2009) were used to deal with multiple measures and multiple comparisons within a single study. In studies including multiple independent subgroups (e.g., separate data is presented for boys and girls in the experimental and control condition), first the summary statistics for the two conditions was recreated, and then this data was used to compute the effect size. In studies with multiple treatment groups, the approach taken depended on the nature of the guidance. If guidance was of the same type (e.g., two prompted conditions were compared against one unprompted control condition), a composite effect size was calculated by averaging the effect sizes obtained in the treatment groups. The effect size variance was computed based on the variance of each effect size as well as their intercorrelation. If different types of guidance were assessed within the same study, combining both effects would go against the purpose of this meta-analysis. Instead, the type of guidance that best matched the study's goals and research questions was selected or, if this proved impossible, one type of guidance was selected at random.

A similar approach was used to resolve dependence that occurred when a study reported multiple scores for a single outcome measure. Examples include studies that assess various inquiry skills to indicate learning activities, or separate posttest scores for conceptual and procedural knowledge. As with multiple treatment groups, the preferred solution was to compute a combined effect size and

calculate its variance by taking into account the correlation among the separate scores. When this correlation was not reported, the most relevant score was selected or, as the second-best alternative, one score was chosen at random.

Data Analysis

As not every study included all three outcome measures (see Table 2), the shifting-unit-of-analysis approach (Cooper, 1998) was used to preserve as much of the data as possible. This resulted in three separate analyses, one for each outcome measure, which were conducted using an identical approach. All main analyses were performed using the SPSS macros provided by Wilson (2005). All analyses used the random effects model because studies examining different types of guidance and divergent age groups are unlikely to share the same true effect size. Under the random effects model, the observed effect size Y_i is defined as $Y_i = \mu + \xi_i + \epsilon_i$, meaning that the observed effect size for any study is determined by the grand mean (μ), the deviation of the study's true effect size from the grand mean (ξ_i), and the sampling error ϵ_i . The grand mean is estimated by

$$\hat{\mu} = \frac{\sum_{i=1}^k W_i Y_i}{\sum_{i=1}^k W_i} \tag{1}$$

where the weight assigned to a particular study (W_i) is given by $1/(\text{Var}(Y_i) + \tau^2)$, with $\text{Var}(Y_i)$ being the within-study variation and τ^2 the between-study variation. The parameter τ^2 is estimated by $Q - (k - 1)/c$, where Q is the between-study heterogeneity statistic, k is the number of studies, and c is calculated from the studies' weight as

$$c = \sum_{i=1}^k W_i - \frac{\sum_{i=1}^k W_i^2}{\sum_{i=1}^k W_i} \tag{2}$$

To test this meta-analysis' first hypothesis, the summary effect of inquiry learning guidance was analyzed for each outcome measure. Inverse variance weights were used to assign more weight to effect sizes from studies with larger samples. The significance of the summary effect was determined by z tests; the Rosenthal's (1979) fail-safe N was calculated to indicate how many additional studies with a zero mean effect size would be needed to undo a significant overall effect. Next, Q tests based on analysis of variance were performed to examine whether the basic study features moderated the findings. Analog to a one-way ANOVA, these tests indicated whether the observed between-study variance (Q_b) was statistically significant; the I^2 statistic was computed to indicate how much of this variance reflects true score variation. To adjust for multiple comparisons, Hochberg's (1988) step-up procedure was used to control the familywise Type I error rate at level $\alpha = .05$.

The Hochberg-corrected Q_b was also used to test the second hypothesis, which predicted that the overall mean effect size was related to the type of guidance. When homogeneity of effect sizes across the six types of guidance was rejected,

TABLE 3*Summary of effect sizes for learning activities*

	<i>k</i>	<i>d</i>	<i>SE</i>	95% CI	Q_b	df_q	p_q	I^2
Overall	20	0.66	.11	[0.44, 0.88]				
Outcome focus					1.49	1	.223	32.98
Inquiry skills	16	0.73	.12	[0.48, 0.97]				
Regulative skills	4	0.39	.25	[-0.10, 0.87]				
Publication type					0.94	2	.624	0.00
Journal article (IF ≥ 1.5)	14	0.61	.13	[0.35, 0.88]				
Journal article (IF < 1.5)	5	0.70	.22	[0.26, 1.16]				
Dissertation	1	1.13	.52	[0.10, 2.15]				
Study design					0.00	1	.973	0.00
Randomized experiment	12	0.66	.15	[0.36, 0.96]				
Quasi-experiment	8	0.66	.17	[0.32, 1.01]				
Duration					0.80	1	.372	0.00
Single session	15	0.71	.12	[0.46, 0.95]				
Multiple sessions	5	0.46	.25	[-0.03, 0.95]				
Type of guidance					4.01	5	.547	0.00
Process constraints	4	0.90	.23	[0.45, 1.35]				
Status overviews	2	0.29	.32	[-0.33, 0.91]				
Prompts	2	0.51	.29	[-0.05, 1.07]				
Heuristics	1	0.24	.41	[-0.56, 1.04]				
Scaffolds	6	0.77	.23	[0.31, 1.22]				
Explanations	5	0.70	.19	[0.34, 1.07]				

Note. $N = 2,374$.

the planned comparison procedures described in Hedges and Pigott (2004) were used to reveal which types of guidance differed significantly from one another. The Q tests' within-group variance (Q_w) was used to test the third hypothesis that more specific types of guidance are more effective for younger learners and vice versa. The Q_w homogeneity statistic indicated whether it is reasonable to assume that studies investigating the same type of guidance share a common effect size. With the type of guidance for which this assumption was disproved, planned contrasts were performed to examine whether the heterogeneity was attributable to the learners' age group.

Results

Learning Activities

Measures of learning activities were included in 20 of the 72 studies. The summary statistics presented in Table 3 indicate that there was a significant overall effect of guidance in that guided inquiry learning led to a more proficient use of inquiry skills than did unguided inquiry learning, $z = 5.85$, $p < .001$. The d statistic and its confidence interval show that the true effect size is likely to fall in the

range of 0.44 to 0.88, which according to Cohen (1988) can be considered a medium to large effect. Rosenthal's fail-safe procedure further showed that an additional 1,419 null-effect studies would be needed to bring the p level beyond the .05 threshold of significance.

Moderator analysis was performed to examine whether the between-study variation in effect sizes was attributable to the basic study features listed in Table 3. Results showed that this was not the case, as neither outcome focus nor publication type, study design, and duration moderated the findings. Moderator analysis also sought to determine whether the heterogeneity of effect sizes was due to the type of guidance. Again no significant moderation effect was found, which means that the overall effect of guidance on learning activities did not depend on the specificity of the assistance students received.

However, there was significant within-group variation in the overall effect of guidance, $Q_w(14) = 24.50, p = .040$, which could point to possible differential effects due to the learners' age. Indeed, some age-related differences were found for two types of guidance, and the direction of effect was consistent with hypotheses. Process constraints, which is the least directive type of guidance, appeared to be more beneficial for adolescents ($d = 0.94$) than for children ($d = 0.78$). Scaffolds, in contrast, provide rather specific guidance and were found to be more effective for teenagers ($d = 3.62$) compared with adolescents ($d = 0.70$). Statistical analysis nevertheless revealed that none of these differences was statistically significant, $Q_w < 9.77, p > .062$.

Performance Success

A total of 17 studies analyzed the products learners created during the inquiry and used this data to assess the impact of guidance on performance success. The overall mean effect sizes reported in Table 4 indicate that learners who received guidance outperformed their unguided counterparts by more than half a standard deviation, $z = 7.30, p < .001$, which constitutes a medium to large effect (Cohen, 1988). Rosenthal's fail-safe N estimated that 753 unpublished studies with zero mean effect would be required to make the obtained overall effect statistically nonsignificant.

Moderator analysis (with Hochberg's multiple testing correction) showed that the variation between the results of the 17 studies was independent of publication type and duration. However, a significant moderating effect was found for study design. As shown in Table 4, the mean effect size of randomized experiments was almost twice as high as that of quasi-experiments. The I^2 statistic further indicated that more than 80% of the between-study variance in effect size reflects true score variation, which can be considered a high amount of variance according to the benchmarks proposed by Higgins, Thompson, Deeks, and Altman (2003).

The Hochberg-corrected Q_b for type of guidance was also significant, and the magnitude of I^2 indicated that a moderate 65% of the observed variance represents true between-studies variation (Higgins et al., 2003). Consistent with expectations, larger effect sizes were associated with more specific guidance, $r = .53, p = .030$. Planned contrasts further showed that explanations were more effective than all less specific types of guidance combined, $z = 1.84, p = .033$. Along the same lines, no significant difference was found for scaffolds, $z = 1.45, p = .073$, but

TABLE 4*Summary of effect sizes for performance success*

	<i>k</i>	<i>d</i>	<i>SE</i>	95% CI	<i>Q_b</i>	<i>df_q</i>	<i>p_q</i>	<i>I²</i>
Overall	17	0.71	.10	[0.52, 0.90]				
Publication type					2.19	1	.139	54.33
Journal article (IF ≥ 1.5)	10	0.59	.12	[0.36, 0.82]				
Journal article (IF < 1.5)	7	0.86	.14	[0.59, 1.12]				
Study design					5.91	1	.015	83.08
Randomized experiment	8	0.94	.13	[0.68, 1.19]				
Quasi-experiment	9	0.54	.10	[0.34, 0.74]				
Duration					0.74	1	.390	0.00
Single session	9	0.64	.12	[0.40, 0.88]				
Multiple sessions	8	0.80	.14	[0.52, 1.08]				
Type of guidance					14.34	5	.014	65.13
Process constraints	4	0.71	.14	[0.43, 0.99]				
Status overviews	2	0.22	.20	[-0.18, 0.61]				
Prompts	3	0.50	.14	[0.23, 0.76]				
Heuristics	1	1.17	.34	[0.50, 1.84]				
Scaffolds	6	0.80	.13	[0.55, 1.05]				
Explanations	1	1.45	.37	[0.72, 2.18]				

Note. *N* = 1,019.

heuristics were significantly more effective than their less specific alternatives, $z = 1.72$, $p = .043$. Prompts were as effective as status overviews and process constraints combined, $z = 0.25$, $p = .401$, and status overviews were less effective than process constraints, $z = 2.01$, $p = .023$.

The within-group test of homogeneity was not significant, $Q_w(11) = 15.42$, $p = .164$. This finding means that a particular type of guidance has a similar positive effect on performance in children, teenagers, and adolescents.

Learning Outcomes

The majority of the studies included in this meta-analysis addressed the effects of guidance on learning outcomes. The overall mean effect size of these 60 studies indicates that the presence of a particular type of guidance had a significant positive effect on learning outcomes, $z = 7.83$, $p < .001$ (see Table 5). The magnitude of the summary effect was approximately half a standard deviation, which is a medium effect according to Cohen's (1988) benchmarks. Rosenthal's fail-safe procedure indicated that 6,379 zero-effect studies should be added to the sample to bring the mean effect size down to a statistically nonsignificant level.

The Q_b tests reported in Table 5 indicate that two basic study features might account for the observed variation in effect sizes. Publication type tended to moderate the findings such that journal articles contained higher effect sizes than conference papers and dissertations combined, and journal impact factor was positively correlated with effect size, $r = .31$, $p = .049$. However, these effects

TABLE 5
Summary of effect sizes for learning outcomes

	<i>k</i>	<i>d</i>	<i>SE</i>	95% CI	<i>Q_b</i>	<i>df_q</i>	<i>p_q</i>	<i>I²</i>
Overall	60	0.50	.06	[0.37, 0.62]				
Outcome focus					9.43	1	.002	89.39
Inquiry skills	18	0.78	.11	[0.56, 1.00]				
Domain knowledge	42	0.37	.07	[0.23, 0.52]				
Publication type					8.42	3	.036	64.37
Journal article (IF ≥ 1.5)	32	0.65	.09	[0.48, 0.82]				
Journal article (IF < 1.5)	16	0.38	.12	[0.14, 0.62]				
Conference paper	9	0.18	.15	[-0.11, 0.48]				
Dissertation	3	0.45	.28	[-0.10, 1.01]				
Domain					1.13	1	.288	11.50
Math	5	0.28	.22	[-0.15, 0.71]				
Science	55	0.52	.07	[0.39, 0.66]				
Study design					0.79	1	.376	0.00
Randomized experiment	33	0.55	.09	[0.38, 0.73]				
Quasi-experiment	27	0.44	.10	[0.25, 0.62]				
Duration					2.23	1	.136	55.16
Single session	36	0.42	.08	[0.25, 0.58]				
Multiple sessions	24	0.61	.10	[0.42, 0.81]				
Type of guidance					4.05	5	.542	0.00
Process constraints	4	0.52	.25	[0.03, 1.01]				
Status overviews	2	0.61	.35	[-0.09, 1.30]				
Prompts	8	0.55	.18	[0.19, 0.91]				
Heuristics	2	0.22	.34	[-0.44, 0.88]				
Scaffolds	19	0.34	.12	[0.11, 0.56]				
Explanations	25	0.61	.10	[0.42, 0.80]				

Note. *N* = 5,629.

became statistically nonsignificant after Hochberg correction for multiple comparisons was applied. The significant moderating effect of outcome focus remained, which means that the effect sizes of studies assessing inquiry skills was more than twice as high as the effect sizes found in studies assessing domain knowledge. The benchmarks for *I²* proposed by Higgins et al. (2003) imply that a high amount of the between-studies variance was real rather than spurious.

The type of guidance had no significant moderation effect, meaning that all six types of guidance, regardless of their specificity, were equally effective in promoting learning outcomes. Analysis of the within-group variance further showed that this effectiveness was independent of the learners' age, $Q_w(54) = 62.73, p = .194$: children, teenagers, and adolescents all benefited to the same extent from a particular type of guidance.

Discussion

This article quantitatively synthesized the results of 72 studies examining the effectiveness of guidance in inquiry-based teaching and learning. The obtained findings confirm the first hypothesis that guidance has a significant positive effect on inquiry learning activities, performance success, and learning outcomes. These results are consistent with recent research integrations comparing guided and unguided inquiry learning with other instructional methods (Alfieri et al., 2011; Furtak et al., 2012) and corroborate the conclusion of related meta-analyses that addressed the issue of guidance from a slightly different perspective. The one by Carolan et al. (2014) focused on simulation-based inquiry learning in adult education and imposed no date restrictions on the included studies; the resulting set of 31 studies contained none of the empirical works included in the present meta-analysis. D'Angelo et al. (2014) limited their search of the literature on simulation-based inquiry learning to the same time period as the present meta-analysis, but included only five of the studies listed in Table 2 in their meta-analysis. The present meta-analysis thus adds to the research by showing that the benefits of guidance generalize beyond simulation-based inquiry learning, are applicable to different age categories, and extend to learning activities and performance success.

Moderator analysis did not lead to a uniform conclusion as to whether the variation in effect sizes was attributable to basic study features. Two characteristics were found to moderate the findings, but their effects were inconsistent across the three meta-analyses. The overall effect of guidance on performance success was moderated by the studies' research design. Randomized experiments yielded significantly higher effect sizes than quasi-experiments, which suggest that methodological quality could be one explanation for the divergence of effect sizes. Regarding learning outcomes, the moderating effect of outcome focus indicates that guidance has a larger impact on the development of inquiry skills than on the acquisition of domain knowledge. One plausible explanation is that the guidance included in these meta-analyses was exclusively geared toward inquiry skills; inducing knowledge by performing these skills is possible but at the same time susceptible to errors not necessarily addressed by the guidance.

Of further interest is that certain study features did *not* moderate the effect sizes. One of these concerned the domain of the studies. A mere 7% of the included studies were in math education but unlike D'Angelo et al. (2014), where math education was equally underrepresented, the domain did not moderate the effectiveness of guidance. The studies' duration did not affect the effectiveness of guidance either. This finding challenges the postulation that long-term intervention studies are needed to establish the true effect of guided inquiry learning. Similar to D'Angelo et al., the present findings indicate that short-term studies reveal its actual impact just as well. Publication type did not moderate the findings either. This result contradicts the tacit belief among educational scholars that studies with large effects appear more often in esteemed outlets than studies with medium or small effects.

The second research question concerned the relative effectiveness of different types of guidance. More specific guidance such as scaffolds and explanations was expected to yield larger effect sizes than less specific support such as process

constraints or prompts. This hypothesis was partially supported by the results. The overall mean effect size for learning activities and learning outcomes was independent of the type of guidance, but a moderator effect was found in performance success. The rank correlation confirmed that larger effect sizes were associated with more specific types of guidance. Results of the planned contrasts, although generally supportive, do not allow for any definitive conclusion because several comparisons included too few effect sizes from which to make valid generalizations. Still, the overall conclusion seems to be that learners perform better during an inquiry (i.e., create better products to exhibit their domain knowledge) when supported by more specific forms of guidance.

This conclusion raises the question why the type of guidance did not affect learning outcomes, which in 42 of the 60 studies concerned domain knowledge. Part of the answer lies in the lower effect sizes for domain knowledge compared with inquiry skills. In keeping with the aforementioned explanation, offering more specific guidance has an immediate effect on the domain knowledge articulated during the inquiry, but these initial differences fade away quickly after the instructional context is withdrawn. This reflects the incommensurability between learning and performance: instructional interventions designed to increase performance during the learning task are often less effective in increasing learning in the long term (Kapur & Rummel, 2012). The guidance included in this meta-analysis was at the skill level and might therefore be less appropriate to promote learning outcomes.

In light of this explanation one would expect the type of guidance to moderate learning activities. This was not the case, but the considerable within-group variance points to a differential effect that could be due to the learners' age. This effect was most apparent for process constraints and scaffolds, and consistent with the assumption that more specific types of guidance are more beneficial for younger learners. Still, the number of studies in this analysis was rather low which does not allow for any definitive conclusions; other factors not included in the moderator analysis could have caused the within-group variation as well. An alternative explanation for the absence of any consistent age-related differences could be that empirical researchers accommodate the inquiry tasks used in their studies to the capabilities of the age group. If so, young learners engage in less demanding tasks than older learners, which cancels out possible age-related differences in proficiency and hence causes various age groups to benefit as much from more or less specific types of guidance. This possibility would obviously complicate the assessment of the effectiveness of guidance and points to a need for more cross-sectional research in which the same task is used with participants from different age groups (cf. Penner & Klahr, 1996; Schauble, 1996; Tschirgi, 1980; Veenman et al., 2004) randomly assigned to either a guided or unguided inquiry condition.

Implications for Theory and Research

The insights gained through these meta-analyses point to the need to reconsider the assumptions accompanying the typology of guidance. This framework grew out of a recent review of the inquiry learning literature (T. De Jong & Lazonder, 2014) and was applied for the first time in this meta-analysis. The framework proved useful to classify the guidance used in previous studies on a

single dimension, but the assumed match between the types of guidance and learners' prior knowledge and skills was contradicted by the current findings. Despite the moderating effect in performance success, learning activities and learning outcomes were equally enhanced by each type of guidance. These results suggest that less specific forms of guidance are already quite useful to young learners with low inquiry skills, and that older, more experienced learners do benefit from specific types of guidance such as scaffolds and explanations. These new insights should be addressed in subsequent writings to further improve the typology.

On the other hand, the present findings shed new light on the conditions of effective minimally guided instructional approaches such as inquiry-based and problem-based learning. Underlying Kirschner et al.'s (2006) plea for high instructional guidance is the tacit assumption that more specific guidance leads to greater learning. Kirschner et al. advocated for worked examples and process worksheets as effective methods of guided learning which both have been successfully applied to facilitate inquiry-based learning (e.g., De Vries et al., 2008; Mulder, Lazonder, & De Jong, 2014). The results of the meta-analyses confirm that even though worked examples and process worksheets differ in specificity—they would be classified as explanations and heuristics, respectively—both would yield similar positive effects on learning and performance. This suggests that “strong guidance” does not necessarily mean “specific guidance;” other dimensions such as the frequency and duration of the guidance might be more imperative to help reduce the demands of inquiry-based learning on working memory. Future research should examine whether and how these properties contribute to the effectiveness of guidance to support inquiry learning activities and outcomes.

Future research should also address the issue of how multiple types of guidance are best combined. The present meta-analysis focused on the effects of individual types of guidance. Although such controlled investigations warrant experimental rigor, inquiry learning practices often incorporate various forms of help and assistance. For example, teachers may start the inquiry with a short benchmark lesson; give hints, feedback, and explanations during the inquiry; and discuss the learners' performance after the inquiry. The questions as to how different types of guidance are best combined, and whether such a combination is more effective than offering a single type of guidance have received minimal attention in empirical investigations. The orchestration of guidance therefore merits attention in future studies.

The present findings do not point to restrictions regarding the duration of future studies on guided inquiry learning. This implication contradicts the common concern that inquiry learning interventions are often studied for a too short period of time (e.g., T. De Jong & Lazonder, 2014; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995). The current meta-analysis proved otherwise and imply that the true effects of guidance can be revealed in short-term studies. However, this is not to say that adequate guidance causes all inquiry skills to develop at the same pace. More demanding skills such as evidence evaluation inherently require more time to develop than more straightforward skills and do not always reach maturity (e.g., Zimmerman, 2007). Guidance facilitates the development process and effects observed in the short term remain when the duration of the inquiry is

extended. Yet guidance in and of itself does not serve as catalyst to boost the pace of development.

Implications for Teaching and Learning

The results of these meta-analyses point to several implications for practice. First and foremost, inquiry-based teaching practices should employ guidance to assist learners in accomplishing the task and learn from the activity. The need for guidance applies to short-term inquiries that are performed as part of a single lesson as well as to more comprehensive inquiry units and projects that span multiple lessons. So whenever learners act like scientists, their teacher should provide them with adequate guidance.

Adequate guidance is not the same as highly specific guidance. Too much guidance inevitably challenges the inherent nature of the inquiry process, and the present findings indicate that less specific forms of guidance lead to comparable learning activities and outcomes as more specific guidance. This enables teachers to create guided learning environments that give learners enough freedom to examine a topic or perform a task on their own. For example, math teachers do not need to direct students through every step of the mathematical modeling process: giving them a heuristic that conveys the gist of each phase will generally suffice. Likewise, elementary science teachers can teach children the basics of experimental design prior to an inquiry, but the use of these skills is equally enhanced by introducing the variables under investigation one at a time. The deployment of such learning environments would help meet the need for more authentic school inquiry tasks that reflect the core attributes of scientific practice (Chinn & Malhotra, 2002b). Engaging in these tasks enables even the youngest learners to learn to understand and appreciate the methods and epistemology of scientific investigation which might eventually cause them to pursue a career in science or mathematics.

However, highly specific guidance is necessary when teachers want students to maximize their performance, for instance when a science project serves as a showcase for parents or when student products are submitted to a national contest. The moderating effect of guidance found in the second meta-analysis indicates that performance success increases more when learners receive more specific guidance. Although the tentative nature of the conclusions imply that this recommendation may not apply to all six types of guidance, the mean effect sizes seem to suggest that explaining a yet unmastered inquiry skill in full enhances performance more than does any of the other, less specific forms of guidance.

Finally, teachers do not have to take the age of their students into account in selecting a particular type of guidance. The conclusions and recommendations presented thus far all apply to children, teenagers, and adolescents. In absence of any consistent age-related differences, teachers can base their choice of guidance on other factors such as the learners' topical knowledge or familiarity with the inquiry skills, and the teacher–student ratio. As a result, elementary school teachers may decide to give children a simple prompt instead of extensive explanations, and teachers in middle and high school classes can support their students through specific directions or a mini lesson at the start of an inquiry. The decision

to increase or decrease the specificity of the guidance learners receive rests on factors other than the ones included in these meta-analyses.

Limitations

Any statistical research integration inherently has limitations, and the present meta-analysis is no exception. A common problem is that some well-designed empirical studies tend to report incomplete statistical information and for that reason six studies had to be excluded from the present meta-analysis. In addition to reducing statistical power, this discard also decreased the number of second-tier journal articles and conference papers which were already underrepresented in the sample. This in turn muddies the waters concerning issues of publication bias in that informative studies had to be discarded for “technical” reasons rather than nonsignificant findings.

A related limitation concerns the relatively low number of primary studies in the meta-analyses of learning activities and performance success. This complicated the analysis of possible specificity effects of guidance in performance success and might explain why the age-related differences in learning activities failed to show. The limited set of studies also caused a more lenient stance toward handling multiple comparisons. Many dependencies were controlled for, for example by extracting only one dependent variable and one treatment group from each study, and examining their effects in a separate meta-analysis for each outcome measure, but some studies were included in two or three meta-analyses. Although the amount of overlap was acceptable (72% of the studies was included in just one meta-analysis), there was no complete independence.

Another issue concerns the conduct of multiple moderator analyses on the same, small set of studies. To control the familywise Type I error rate, omnibus tests of heterogeneity preceded planned comparisons, and Hochberg’s step-up procedure was used where appropriate to ensure that the overall likelihood of a Type I error would not exceed 5%. Although Hochberg’s procedure is less well-known than, for example, a Bonferroni correction, it is also less conservative in that it maximizes the possibility to detect true effects. Still, any adjustment for multiple comparisons increases the probability of producing false negatives, and hence reduces statistical power. This was perhaps most apparent in the moderator analysis of learning outcomes, where the effect of the type of publication turned statistically nonsignificant after adjustment. A practical yet improper solution would be to reduce the number of moderators (even after the fact) so that maximum statistical significance is maintained. A more sincere solution, which was used here, is to present enough information for anyone to evaluate the data in context.

Finally, the scope of this meta-analysis was limited to disciplines subsumed under math and science education. In view of the recent emphasis on STEM education, it might seem like a missed opportunity to exclude studies in technology and engineering education. Yet this was a deliberate choice made for practical reasons because the omitted disciplines are not generally taught in K-12 classes. This decision nevertheless restricts the generalization of the findings to STEM education as a whole.

Conclusion

This synthesis of 72 empirical studies demonstrates that guidance is pivotal to successful inquiry-based learning. Learners who are given some kind of guidance act more skillfully during the task, are more successful in obtaining topical information from their investigational practices, and score higher on tests of learning outcomes administered after the inquiry. These benefits are largely independent of the specificity of the guidance: even though performance success tends to increase more when more specific guidance is available, learning activities and learning outcomes improve as much with specific and nonspecific types of guidance. The effectiveness of guidance applies equally to children, teenagers, and adolescents, which offers educational designers and K-12 math and science teachers a wide choice of opportunities to effectively involve learners in forms of inquiry-based learning.

Note

This meta-analysis is part of the research program “Inquiry Learning Support for Children and Adolescents: What Works When and Why?” (411-12-201), which was financed by the Netherlands Organization for Scientific Research (NWO).

References

- References marked with an asterisk indicate studies included in the meta-analysis.
- Abd-El-Khalick, F., BouJaoude, S., Duschl, R., Lederman, N., Mamiok-Naaman, R., Hofstein, A., . . . Tuan, H. (2004). Inquiry in science education: International perspectives. *Science Education*, *88*, 397–419. doi:10.1002/sce.10118
- *Abdullah, S., & Shariff, A. (2008). The effects of inquiry-based computer simulation with cooperative learning on scientific thinking and conceptual understanding of gas laws. *Eurasia Journal of Mathematics, Science & Technology Education*, *4*, 387–398. Retrieved from http://www.ejmste.org/v4n4/EURASIA_v4n4_Abdullah.pdf
- Achieve. (2010). *International science benchmarking report. Taking the lead in science education: Forging next-generation science standards*. Retrieved from <http://www.achieve.org/files/InternationalScienceBenchmarkingReport.pdf>
- Alessi, S. (1995, April). *Dynamic versus static fidelity in a procedural simulation*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology*, *103*, 1–18. doi:10.1037/a0021017
- Ausubel, D. P. (1962). Learning by discovery. *Educational Leadership*, *20*, 113–117.
- *Baroody, A. J., Eiland, M. D., Purpura, D. J., & Reid, E. E. (2012). Fostering at-risk kindergarten children’s number sense. *Cognition and Instruction*, *30*, 435–470. doi: 10.1080/07370008.2012.720152
- *Baroody, A. J., Eiland, M. D., Purpura, D. J., & Reid, E. E. (2013). Can computer-assisted discovery learning foster first graders’ fluency with the most basic addition combinations? *American Educational Research Journal*, *50*, 533–573. doi:10.3102/0002831212473349
- Bittinger, M. L. (1968). A review of discovery. *The Mathematics Teacher*, *61*, 140–146.

- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: Wiley.
- Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, *31*, 21–32.
- Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York, NY: Wiley.
- Burns, B. D., & Vollmeyer, R. (2002). Goal specificity effects on hypothesis testing in problem solving. *Quarterly Journal of Experimental Psychology*, *55A*, 241–261. doi:10.1080/02724980143000262
- Cañadas, M. C., Deulofeu, J., Figueiras, L., Reid, D., & Yevdokimov, O. (2007). The conjecturing process: Perspectives in theory and implications in practice. *Journal of Teaching and Learning*, *5*, 55–72. Retrieved from <http://ojs.uwindsor.ca/ojs/leddy/index.php/JTL/article/view/82>
- Carolan, T. F., Hutchins, S. D., Wickens, C. D., & Cumming, J. M. (2014). Costs and benefits of more learner freedom: Meta-analyses of exploratory and learner control training methods. *Human Factors*, *56*, 999–1014. doi:10.1177/0018720813517710
- *Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, *70*, 1098–1120. doi:10.1111/1467-8624.00081
- *Chin, D. B., Dohmen, I. M., & Schwartz, D. L. (2013). Young children can learn scientific reasoning with teachable agents. *IEEE Transactions on Learning Technologies*, *6*, 248–257. doi:10.1109/Tlt.2013.24
- Chinn, C. A., & Malhotra, B. A. (2002a). Children's responses to anomalous scientific data: How is conceptual change impeded? *Journal of Educational Psychology*, *94*, 327–343. doi:10.1037/0022-0663.94.2.327
- Chinn, C. A., & Malhotra, B. A. (2002b). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*, 175–218. doi:10.1002/sce.10001
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cooper, H. (1998). *Synthesizing research: A guide for literature reviews*. Thousand Oaks, CA: Sage.
- D'Angelo, C., Rutstein, D., Harris, C., Bernard, R., Borokhovski, E., & Haertel, G. (2014). *Simulations for STEM learning: Systematic review and meta-analysis*. Menlo Park, CA: SRI International.
- *Davis, E. A. (2000). Scaffolding students' knowledge integration: Prompts for reflection in KIE. *International Journal of Science Education*, *22*, 819–837. doi:10.1080/095006900412293
- *De Jong, K., & Brand-Gruwel, S. (2008). Stimuleren van ontdekkend leren met behulp van proceswerkbladen: Effecten op taakprestaties [Stimulating inquiry learning by means of process worksheets: Effects on pupils' task performance]. *Pedagogische Studiën*, *85*, 451–475.
- De Jong, T. (2006). Scaffolds for scientific discovery learning. In J. Elen & R. E. Clark (Eds.), *Handling complexity in learning environments: Theory and research* (pp. 107–128). London, England: Elsevier.
- De Jong, T., & Lazonder, A. W. (2014). The guided discovery learning principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 371–390). New York, NY: Cambridge University Press.

- De Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science, 340*, 305–308. doi:10.1126/science.1230579
- *De Jong, T., Martin, E., Zamarro, J. M., Esquembre, F., Swaak, J., & Van Joolingen, W. R. (1999). The integration of computer simulation and learning support: An example from the physics domain of collisions. *Journal of Research in Science Teaching, 36*, 597–615. doi:10.1002/(Sici)1098-2736(199905)36:5<597::Aid-Tea6>3.0.Co;2-6
- De Jong, T., Soteriou, S., & Gillet, D. (2014). Innovations in STEM education: The Go-Lab federation of online labs. *Smart Learning Environments, 1*(3). doi:10.1186/s40561-014-0003-6
- De Jong, T., & Van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*, 179–202. doi:10.3102/00346543068002179
- De Vries, B., Van der Meij, H., & Lazonder, A. W. (2008). Supporting reflective web searching in elementary schools. *Computers in Human Behavior, 24*, 649–665. doi:10.1016/j.chb.2007.01.021
- *Dean, D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education, 91*, 384–397. doi:10.1002/Sc.20194
- *Dejonckheere, P. J. N., Van de Keere, K., & Mestdagh, N. (2009). Training the scientific thinking circle in pre- and primary school children. *Journal of Educational Research, 103*, 1–16. doi:10.1080/00220670903228595
- *Dejonckheere, P. J. N., Van de Keere, K., & Tallir, I. (2011). Are fourth and fifth grade children better scientists through metacognitive learning? *Electronic Journal of Research in Educational Psychology, 9*, 133–156. Retrieved from <http://investigacion-psicopedagogica.org/revista>
- *Dyke, G., Adamson, D., Howley, I., & Rose, C. P. (2013). Enhancing scientific reasoning and discussion with conversational agents. *IEEE Transactions on Learning Technologies, 6*, 240–247. doi:10.1109/Tlt.2013.25
- *Eckhardt, M., Urhahne, D., Conrad, O., & Harms, U. (2013). How effective is instructional support for learning with computer simulations? *Instructional Science, 41*, 105–124. doi:10.1007/s11251-012-9220-y
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings for the National Academy of Sciences of the United States of America, 111*, 8410–8415. doi:10.1073/pnas.1319030111
- *Fund, Z. (2007). The effects of scaffolded computerized science problem-solving on achievement outcomes: A comparative study of support programs. *Journal of Computer Assisted Learning, 23*, 410–424. doi:10.1111/j.1365-2729.2007.00226.x
- Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research, 82*, 300–329. doi:10.3102/0034654312457206
- Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*, 306–355. doi:0.1016/0010-0285(80)90013-4
- Gijlers, H., & De Jong, T. (2005). The relation between prior knowledge and students' collaborative discovery learning processes. *Journal of Research in Science Teaching, 42*, 264–282. doi:10.1002/tea.20056
- *Gijlers, H., & De Jong, T. (2009). Sharing and confronting propositions in collaborative scientific discovery learning. *Cognition and Instruction, 27*, 239–268. doi:10.1080/07370000903014352

- *Gijlers, H., & De Jong, T. (2013). Using concept maps to facilitate collaborative simulation-based inquiry learning. *Journal of the Learning Sciences*, 22, 340–374. doi:10.1080/10508406.2012.748664
- *Graesser, A. C., Wiley, J., Goldman, S. R., O'Reilly, T., Jeon, M., & McDaniel, B. (2007). SEEK web tutor: Fostering a critical stance while exploring the causes of volcanic eruption. *Metacognition and Learning*, 2, 89–105. doi:10.1007/s11409-007-9013-x
- *Hagemans, M. G., Van der Meij, H., & De Jong, T. (2013). The effects of a concept map-based support tool on simulation-based inquiry learning. *Journal of Educational Psychology*, 105, 1–24. doi:10.1037/A0029433
- *Han, I., & Black, J. B. (2011). Incorporating haptic feedback in simulation for learning physics. *Computers & Education*, 57, 2281–2290. doi:10.1016/j.compedu.2011.06.012
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods*, 9, 426–445. doi:10.1037/1082-989X.9.4.426
- Hermann, G. (1969). Learning by discovery: A critical review of studies. *Journal of Experimental Education*, 38, 58–72.
- Higgins, J., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327, 557–560. doi:0.1136/bmj.327.7414.557
- Hmelo-Silver, C. E., Golan Duncan, R., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42, 99–107. doi:10.1080/00461520701263368
- Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75, 800–802. doi:10.1093/biomet/75.4.800
- *Hogan, K. (1999). Thinking aloud together: A test of an intervention to foster students' collaborative scientific reasoning. *Journal of Research in Science Teaching*, 36, 1085–1109. doi:10.1002/(Sici)1098-2736(199912)36:10<1085::Aid-Tea3>3.0.Co;2-D
- *Huang, K. (2011). *Metaconceptually-enhanced simulation-based inquiry learning: Effects on the 8th grade students' conceptual change and science epistemological beliefs* (Unpublished doctoral dissertation). University of Oklahoma, Norman.
- *Hulshof, C. D., & De Jong, T. (2006). Using just-in-time information to support scientific discovery learning in a computer-based simulation. *Interactive Learning Environments*, 14, 79–94. doi:10.1080/10494820600769171
- *Hushman, C. (2011). *Examining the influence of instructional strategy on student learning and self-efficacy in science* (Unpublished doctoral dissertation). The University of New Mexico, Albuquerque.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York, NY: Basic Books.
- Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2007). Visualization of participation: Does it contribute to successful computer-supported collaborative learning? *Computers & Education*, 49, 1037–1065. doi:10.1016/j.compedu.2006.01.004
- Kanari, Z., & Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching*, 41, 748–769. doi:10.1002/tea.20020

- Kaplan, D. E., & Black, J. B. (2003). Mental models and computer-based scientific inquiry learning: Effects of mechanistic cues on adolescent representation and reasoning about causal systems. *Journal of Science Education and Technology, 12*, 483–493. doi:10.1023/B:JOST.0000006308.01183.85
- Kapur, M., & Rummel, N. (2012). Productive failure in learning from generation and invention activities. *Instructional Science, 40*, 645–650. doi:10.1007/s11251-012-9235-4
- *Keselman, A. (2003). Supporting inquiry learning by promoting normative understanding of multivariable causality. *Journal of Research in Science Teaching, 40*, 898–921. doi:10.1002/Tea.10115
- 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*, 75–86. doi:10.1207/s15326985ep4102_1
- Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes*. Cambridge: MIT Press.
- Klahr, D. (2013). What do we mean? On the importance of not abandoning scientific rigor when talking about science education. *Proceedings of the National Academy of Science of the United States of America, 110*, 14075–14080. doi:10.1073/pnas.1212738110
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*, 1–48. doi:10.1207/s15516709cog1201_1
- Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology, 25*, 111–146. doi:10.1006/cogp.1993.1003
- Klahr, D., & Li, J. (2005). Cognitive research and elementary science instruction: From the laboratory, to the classroom, and back. *Journal of Science Education and Technology, 14*, 217–238. doi:10.1007/s10956-005-4423-5
- Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science, 15*, 661–667. doi:10.1111/j.0956-7976.2004.00737.x
- *Knaggs, C. M., & Schneider, R. M. (2012). Thinking like a scientist: Using vee-maps to understand process and concepts in science. *Research in Science Education, 42*, 609–632. doi:10.1007/s11165-011-9213-x
- *Kolloffel, B., Eysink, T. H. S., & De Jong, T. (2010). The influence of learner-generated domain representations on learning combinatorics and probability theory. *Computers in Human Behavior, 26*, 1–11. doi:10.1016/j.chb.2009.07.008
- Koerber, S., Sodian, B., Kropf, N., Mayer, D., & Schwippert, K. (2011). The development of scientific reasoning in elementary school age: Understanding theories, designing experiments, interpreting data. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 43*, 16–21. doi:10.1026/0049-8637/A000027
- Koerber, S., Sodian, B., Thoermer, C., & Nett, U. (2005). Scientific reasoning in young children: Preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology, 64*, 141–152. doi:10.1024/1421-0185.64.3.141
- Koslowski, B. (2008, July). *Scientific reasoning is not necessarily plagued by confirmation bias*. Paper presented at the 29th meeting of the International Congress of Psychology, Berlin, Germany.
- *Kramarski, B., & Dudai, V. (2009). Group-metacognitive support for online inquiry in mathematics with differential self-questioning. *Journal of Educational Computing Research, 40*, 377–404. doi:10.2190/EC.40.4.a

- Kuhn, D. (2005). *Education for thinking*. Cambridge, MA: Harvard University Press.
- Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. Orlando, FL: Academic Press.
- *Kuhn, D., & Dean, D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science*, *16*, 866–870. doi:10.1111/j.1467-9280.2005.01628.x
- Kuhn, D., & Franklin, S. (2006). The second decade: What develops (and how). In W. Damon, R. M. Lerner, D. Kuhn, & R. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th ed., pp. 953–994). Hoboken, NJ: Wiley.
- Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development*, *60*(4), 1–157 (Serial No. 245). doi:0.2307/1166059
- *Kuhn, D., & Katz, J. (2009). Are self-explanations always beneficial? *Journal of Experimental Child Psychology*, *103*, 386–394. doi:10.1016/j.jecp.2009.03.003
- Kuhn, D., & Phelps, E. (1982). The development of problem-solving strategies. *Advances in Child Development and Behavior*, *17*, 1–44. doi:10.1016/S0065-2407(08)60356-0
- *Kunzing, J., Kempf, J., & Wirth, J. (2013). Enhancing scientific discovery learning through metacognitive support. *Contemporary Educational Psychology*, *38*, 349–360. doi:10.1016/j.cedpsych.2013.07.001
- Lazonder, A. W. (2014). Inquiry learning. In M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 453–464). Berlin, Germany: Springer.
- *Lazonder, A. W., & Egberink, A. (2014). Children's acquisition and use of the control-of-variables strategy: Effects of explicit and implicit instructional guidance. *Instructional Science*, *42*, 291–304. doi:10.1007/s11251-013-9284-3
- Lazonder, A. W., Hagemans, M. G., & De Jong, T. (2010). Offering and discovering domain information in simulation-based inquiry learning. *Learning and Instruction*, *20*, 511–520. doi:10.1016/j.learninstruc.2009.08.001
- *Lazonder, A. W., & Kamp, E. (2012). Bit by bit or all at once? Splitting up the inquiry task to promote children's scientific reasoning. *Learning and Instruction*, *22*, 458–464. doi:10.1016/j.learninstruc.2012.05.005
- *Lederman, J. S., Lederman, N. G., Wickman, P.-O., & Lager-Nyqvist, L. (2007, April). *An international, systematic investigation of the relative effects of inquiry and direct instruction*. Paper presented at the Annual International Conference of the National Association for Research in Science Teaching, New Orleans, LA.
- Lee, C.-Y., & Chen, M.-P. (2009). A computer game as a context for non-routine mathematical problem solving: The effects of type of question prompt and level of prior knowledge. *Computers & Education*, *52*, 530–542. doi:10.1016/j.compedu.2008.10.008
- *Lin, X. D., & Lehman, J. D. (1999). Supporting learning of variable control in a computer-based biology environment: Effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching*, *36*, 837–858. doi:10.1002/(Sici)1098-2736(199909)36:7<837::Aid-Tea6>3.0.Co;2-U
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- *Lorch, R. F., Lorch, E. P., Calderhead, W. J., Dunlap, E. E., Hodell, E. C., & Freer, B. D. (2010). Learning the control of variables strategy in higher and lower achieving

- classrooms: Contributions of explicit instruction and experimentation. *Journal of Educational Psychology*, 102, 90–101. doi:10.1037/A0017972
- *Lorch, R. F., Lorch, E. P., Freer, B. D., Dunlap, E. E., Hodell, E. C., & Calderhead, W. J. (2014). Using valid and invalid experimental designs to teach the control of variables strategy in higher and lower achieving classrooms. *Journal of Educational Psychology*, 106, 18–35. doi:10.1037/a0034375
- *MacGregor, S. K., & Lou, Y. (2005). Web-based learning: How task scaffolding and web site design support knowledge acquisition. *Journal of Research on Technology in Education*, 37, 161–175. doi:10.1080/15391523.2004.10782431
- *Mandrin, P.-A., & Preckel, D. (2009). Effect of similarity-based guided discovery learning on conceptual performance. *School Science and Mathematics*, 109, 133–145. doi:10.1111/j.1949-8594.2009.tb17949.x
- *Manlove, S., Lazonder, A. W., & De Jong, T. (2006). Regulative support for collaborative scientific inquiry learning. *Journal of Computer Assisted Learning*, 22, 87–98. doi:10.1111/j.1365-2729.2006.00162.x
- *Manlove, S., Lazonder, A. W., & De Jong, T. (2007). Software scaffolds to promote regulation during scientific inquiry learning. *Metacognition and Learning*, 2, 141–155. doi:10.1007/s11409-007-9012-y
- *Mannheimer Zydney, J. (2005, April). *Deeping our understanding through multiple perspectives: The effectiveness of scaffolding in a learning environment*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.
- *Marschner, J., Thillmann, H., Wirth, J., & Leutner, D. (2012). Wie lässt sich die Experimentierstrategie-Nutzung fördern? Ein Vergleich verschieden gestalteter Prompts [How can the use of strategies for experimentation be fostered? A comparison of differently designed prompts]. *Zeitschrift für Erziehungswissenschaft*, 15, 77–93. doi:10.1007/s11618-012-0260-5
- *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, 621–634. doi:10.1007/s11251-012-9248-z
- *Milne, C., Plass, J. L., Homer, B. D., Schwartz, R. N., Khan, M., Ching, D., & Chang, Y. K. (2011, April). *Developing narrative scaffolds for use within multimedia chemistry simulations: Challenges and possibilities*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- *Milne, C., Plass, J. L., Homer, B. D., Wang, Y., Jordan, T., Schwartz, R. N., . . . Hayward, E. (2010, April). *Exploring the possibilities for narrative in the use of multimedia simulations for the teaching and learning of chemistry*. Paper presented at the annual meeting of the American Educational Research Association, Denver, CO.
- *Minogue, J., & Borland, D. (2012, March). *Investigating students' ideas about buoyancy and the influence of haptic feedback*. Paper presented at the Annual International Conference of the National Association for Research in Science Teaching, Indianapolis, IN.
- *Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science*, 32, 99–113. doi:10.1023/B:Truc.0000021811.66966.1d
- *Morgan, K., & Brooks, D. W. (2012). Investigating a method of scaffolding student-designed experiments. *Journal of Science Education and Technology*, 21, 513–522. doi:10.1007/s10956-011-9343-y

- *Mulder, Y. G., Lazonder, A. W., & De Jong, T. (2011). Comparing two types of model progression in an inquiry learning environment with modelling facilities. *Learning and Instruction, 21*, 614–624. doi:10.1016/j.learninstruc.2011.01.003
- Mulder, Y. G., Lazonder, A. W., & De Jong, T. (2014). Using heuristic worked examples to promote inquiry-based learning. *Learning and Instruction, 29*, 56–64. doi:10.1016/j.learninstruc.2013.08.001
- *Mulder, Y. G., Lazonder, A. W., De Jong, T., Anjewierden, A., & Bollen, L. (2012). Validating and optimizing the effects of model progression in simulation-based inquiry learning. *Journal of Science Education and Technology, 21*, 722–729. doi:10.1007/s10956-011-9360-x
- National Research Council. (2000). *Inquiry and the national science standards*. Washington, DC: National Academy Press.
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academy Press.
- National Science Foundation. (2000). An introduction to inquiry. In *Inquiry: Thoughts, views, and strategies for the K-5 classroom* (Vol. 2, pp. 1–5). Washington, DC: Author.
- *Neber, H., & Anton, M. A. (2008). Promoting pre-experimental epistemic activities in chemistry education: Focusing on the role of students' epistemic questions. *International Journal of Science Education, 30*, 1801–1821. doi:10.1080/09500690701579546
- Njoo, M., & De Jong, T. (1993). Exploratory learning with a computer simulation for control theory: Learning processes and instructional support. *Journal of Research in Science Teaching, 30*, 821–844. doi:10.1002/tea.3660300803
- Penner, D. E., & Klahr, D. (1996). The interaction of domain-specific knowledge and domain-general discovery strategies: A study with sinking objects. *Child Development, 67*, 2709–2727. doi:10.1111/1467-8624.ep9706244829
- Piekny, J., Gruber, D., & Maehler, C. (2014). The development of experimentation and evidence evaluation skills at preschool age. *International Journal of Science Education, 36*, 334–354. doi:10.1080/09500693.2013.776192
- Piekny, J., & Maehler, C. (2013). Scientific reasoning in early and middle childhood: The development of domain-general evidence evaluation, experimentation, and hypothesis generation skills. *British Journal of Developmental Psychology, 31*, 153–179. doi:10.1111/j.2044-835X.2012.02082.x
- *Plass, J. L., Homer, B. D., Schwartz, R. N., Milne, C. E., Jordan, T., & Yavner, S. (2013, April). *What you see is what you get: Using visual scaffolds in multimedia simulations*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Puntambekar, S., & Hubscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist, 40*, 1–12. doi:10.1207/s15326985ep4001_1
- Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., . . . Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences, 13*, 337–386. doi:10.1207/s15327809jls1303_4
- *Raes, A., Schellens, T., De Wever, B., & Vanderhoven, E. (2012). Scaffolding information problem solving in web-based collaborative inquiry learning. *Computers & Education, 59*, 82–94. doi:10.1016/j.compedu.2011.11.010

- *Reid, D. J., Zhang, J., & Chen, Q. (2003). Supporting scientific discovery learning in a simulation environment. *Journal of Computer Assisted Learning, 19*, 9–20. doi:10.1046/j.0266-4909.2003.00002.x
- Rey, G. D. (2011). Reset button and instructional advice in computer simulations. *European Psychologist, 16*, 58–67. doi:1027/1016-9040/a000042
- Rieber, L. P., & Parmley, M. W. (1995). To teach or not to teach? Comparing the use of computer-based simulations in deductive versus inductive approaches to learning with adults in science. *Journal of Educational Computing Research, 13*, 359–374. doi:10.2190/M8VX-68BC-1TU2-B6DV
- *Roll, I., Holmes, N. G., Day, J., & Bonn, D. (2012). Evaluating metacognitive scaffolding in guided invention activities. *Instructional Science, 40*, 691–710. doi:10.1007/s11251-012-9208-7
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*, 638–641. doi:10.1037/0033-2909.86.3.638
- *Saab, N., & Van Joolingen, W. R. (2005). Supporting collaborative discovery learning by presenting a tool. In T. Koschmann, D. Suthers, & T.-W. Chan (Eds.), *CSCL 2005: The Next 10 Years, Proceedings of the International Conference on Computer Supported Collaborative Learning 2005* (pp. 541–546). Mahwah, NJ: Erlbaum.
- Samarapungavan, A., Mantzicopoulos, P., & Patrick, H. (2008). Learning science through inquiry in kindergarten. *Science Education, 92*, 868–908. doi:10.1002/sce.20275
- Sao Pedro, M. A., Gobert, J. D., Heffernan, N. T., & Beck, J. E. (2009). Comparing pedagogical approaches for teaching the Control of Variables Strategy. In N. A. Van Taatgen & H. Van Rijn (Eds.), *Proceedings of the 31st Annual Meeting of the Cognitive Science Society* (pp. 1294–1299). Austin, TX: Cognitive Science Society.
- *Sao Pedro, M. A., Gobert, J. D., & Sebuwufu, P. (2011, April). *The effects of self-explanations on robust understanding of the control of variables strategy*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32*, 102–119. doi:10.1037/0012-1649.32.1.102
- Schauble, L., Glaser, R., Duschl, R. A., Schulze, S., & John, J. (1995). Students' understanding of the objectives and procedures of experimentation in the science classroom. *Journal of the Learning Sciences, 4*, 131–166. doi:10.1207/s15327809jls0402_1
- Siler, S., Klahr, D., Strand-Cary, M., & Magaro, C. (2009, April). *Adapting an effective lesson plan for a computer-based tutor*. Paper presented at the Thirteenth European Conference for Research on Learning and Instruction, Amsterdam, Netherlands.
- Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development, 62*, 753–766. doi:10.1111/j.1467-8624.1991.tb01567.x
- *Steffensky, M., Langkes, E.-M., Carstensen, C. H., & Nölke, C. (2012). Alltagssituationen und Experimente: Was sind geeignete naturwissenschaftliche Lerngelegenheiten für Kindergartenkinder? Ergebnisse aus dem SNaKE-Projekt [Daily life and experiments: What are the appropriate learning settings for children in kindergarten? Findings from the SNaKE project]. *Zeitschrift für Erziehungswissenschaft, 15*, 37–54. doi:10.1007/s11618-012-0262-3
- Strand-Cary, M., & Klahr, D. (2008). Developing elementary science skills: Instructional effectiveness and path independence. *Cognitive Development, 23*, 488–511. doi:10.1016/j.cogdev.2008.09.005

- Strike, K. A. (1975). The logic of learning by discovery. *Review of Educational Research*, 45, 461–483. doi:10.3102/00346543045003461
- *Swaak, J., Van Joolingen, W. R., & De Jong, T. (1998). Supporting simulation-based learning: The effects of model progression and assignments on definitional and intuitive knowledge. *Learning and Instruction*, 8, 235–252. doi:10.1037/a0021017
- Taba, H. (1963). Learning by discovery: Psychological and educational rationale. *Elementary School Journal*, 63, 308–316. doi:10.1086/460059
- Toth, E. E., Klahr, D., & Chen, Z. (2000). Bridging research and practice: A cognitively-based classroom intervention for teaching experimentation skills to elementary school children. *Cognition and Instruction*, 18, 423–459. doi:10.1207/S1532690XCI1804_1
- Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development*, 51, 1–10. doi:10.1111/1467-8624.ep12325377
- *Van der Meij, J., Van der Meij, H., & Harmsen, R. (2013, August). *Animated pedagogical agents: Do they enhance student motivation and learning in an inquiry learning environment?* Paper presented at the fifteenth European Conference for Research on Learning and Instruction, Munich, Germany.
- Varma, K. (2014). Supporting scientific experimentation and reasoning in young elementary school students. *Journal of Science Education and Technology*, 23, 381–397. doi:10.1007/s10956-013-9470-8
- *Veenman, M. V. J., Elshout, J. J., & Busato, V. V. (1994). Metacognitive mediation in learning with computer-based simulations. *Computers in Human Behavior*, 10, 93–106. doi:10.1016/0747-5632(94)90031-0
- Veenman, M. V. J., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction*, 14, 89–109. doi:10.1016/j.learninstruc.2003.10.004
- Veermans, K., Van Joolingen, W. R., & De Jong, T. (2006). Use of heuristics to facilitate scientific discovery learning in a simulation learning environment in a physics domain. *Interactive Learning Environments*, 28, 341–361. doi:10.1080/09500690500277615
- *Vreman-de Olde, C., & De Jong, T. (2006). Scaffolding learners in designing investigation assignments for a computer simulation. *Journal of Computer Assisted Learning*, 22, 63–73. doi:10.1111/j.1365-2729.2006.00160.x
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- White, B. Y. (1993). ThinkerTools: Causal models, conceptual change, and science education. *Cognition and Instruction*, 10, 1–100. doi:10.1207/s1532690xci1001_1
- *Wichmann, A., & Leutner, D. (2009). Inquiry learning: Multilevel support with respect to inquiry, explanations and regulation during an inquiry cycle. *Zeitschrift für Pädagogische Psychologie*, 23, 117–127. doi:10.1024/1010-0652.23.2.117
- *Wichmann, A., & Timpe, S. (2013, August). *Increasing interaction by offering variable manipulation possibilities in inquiry learning*. Paper presented at the Fifteenth European Conference for Research on Learning and Instruction, Munich, Germany.
- Wilson, D. B. (2005). *Meta-analysis macros for SAS, SPSS, and Stata*. Retrieved from <http://mason.gmu.edu/~dwilsonb/ma.html>
- *Wollenschlager, M., Moller, J., & Harms, U. (2011). Effects of componential feedback on performance in scientific reasoning. *Zeitschrift für Pädagogische Psychologie*, 25, 197–202. doi:10.1024/1010-0652/A000040

- *Zhang, M. (2013). Supporting middle school students' online reading of scientific resources: Moving beyond cursory, fragmented, and opportunistic reading. *Journal of Computer Assisted Learning*, *29*, 138–152. doi:10.1111/j.1365-2729.2012.00478.x
- Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*, 172–223. doi:10.1016/j.dr.2006.12.001
- *Zohar, A., & Ben David, A. (2008). Explicit teaching of meta-strategic knowledge in authentic classroom situations. *Metacognition and Learning*, *3*, 59–82. doi:10.1007/s11409-007-9019-4
- *Zohar, A., & Peled, B. (2008). The effects of explicit teaching of metastrategic knowledge on low- and high-achieving students. *Learning and Instruction*, *18*, 337–353. doi:10.1016/j.learninstruc.2007.07.001

Authors

ARD W. LAZONDER is an adjunct professor of education, Department of Instructional Technology, University of Twente, PO Box 217, 7500 AE Enschede, the Netherlands; e-mail: a.w.lazonder@utwente.nl. His research interests include simulation-based inquiry learning and scientific reasoning in children.

RUTH HARMSSEN is a PhD student at the Department of Teacher Education, Faculty of Behavioral and Social Sciences, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, the Netherlands; e-mail: r.harmsen@rug.nl. She is a former junior researcher at the Department of Instructional Technology, University of Twente. Her current research interests include designing and evaluating stress reduction intervention programs for beginning teachers.