

Rönnebeck, Silke; Bernholt, Sascha; Ropohl, Mathias Jan

Searching for a common ground - a literature review of empirical research on scientific inquiry activities

Studies in science education 52 (2016) 2, S. 161-197



Quellenangabe/ Reference:

Rönnebeck, Silke; Bernholt, Sascha; Ropohl, Mathias Jan: Searching for a common ground - a literature review of empirical research on scientific inquiry activities - In: *Studies in science education* 52 (2016) 2, S. 161-197 - URN: urn:nbn:de:0111-pedocs-126958 - DOI: 10.25656/01:12695

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-126958>

<https://doi.org/10.25656/01:12695>

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz: <http://creativecommons.org/licenses/by/4.0/deed.de> - Sie dürfen das Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anfertigen, solange Sie den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-License: <http://creativecommons.org/licenses/by/4.0/deed.en> - You may copy, distribute and render this document accessible, make adaptations of this work or its contents accessible to the public as long as you attribute the work in the manner specified by the author or licensor.

By using this particular document, you accept the above-stated conditions of use.

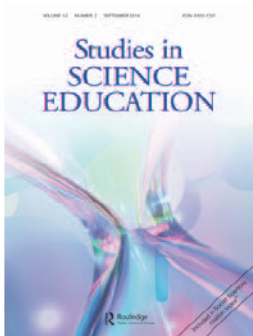


Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft



Searching for a common ground – A literature review of empirical research on scientific inquiry activities

Silke Rönnebeck, Sascha Bernholt & Mathias Ropohl

To cite this article: Silke Rönnebeck, Sascha Bernholt & Mathias Ropohl (2016) Searching for a common ground – A literature review of empirical research on scientific inquiry activities, *Studies in Science Education*, 52:2, 161-197, DOI: [10.1080/03057267.2016.1206351](https://doi.org/10.1080/03057267.2016.1206351)

To link to this article: <http://dx.doi.org/10.1080/03057267.2016.1206351>



© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 01 Aug 2016.



[Submit your article to this journal](#)



Article views: 949



[View related articles](#)



[View Crossmark data](#)

Searching for a common ground – A literature review of empirical research on scientific inquiry activities

Silke Rönnebeck^a, Sascha Bernholt^b and Mathias Ropohl^b

^aDepartment of Educational Research, Leibniz-Institute for Science and Mathematics Education (IPN), Kiel University, Kiel, Germany; ^bDepartment of Chemistry Education, Leibniz-Institute for Science and Mathematics Education (IPN), Kiel University, Kiel, Germany

ABSTRACT

Despite the importance of scientific inquiry in science education, researchers and educators disagree considerably regarding what features define this instructional approach. While a large body of literature addresses theoretical considerations, numerous empirical studies investigate scientific inquiry on quite different levels of detail and also on different theoretical grounds. Here, only little systematic research has analysed the different conceptualisations and usages of the overarching construct of scientific inquiry in detail. To close this gap, a review of the research literature on scientific inquiry was conducted based on a widespread approach to defining scientific inquiry as activities that students engage in. The main goal is to provide a systematic overview about the range and spectrum of definitions and operationalisations used with regard to single activities of the inquiry process in empirical studies. The findings from the review first and foremost illustrate the variability in the ways these activities have been operationalised and implemented. For each activity, studies differ significantly not only with respect to the focus, explicitness and comprehensiveness of their operationalisations but also with regard to the consistency of their implementation in the form of instructional or interventional components in the study and/or in the focus of the assessment of student performance. This has significant implications regarding the validity and comparability of results obtained in different studies, e.g. in the context of discussions concerning the effectiveness of inquiry-based instruction. In addition, the interrelation between scientific inquiry, scientific knowledge and the nature of science seems to be underexplored. The conclusions make the case for further theoretical work as well as empirical research.


KEYWORDS

Scientific inquiry; students' activities; definitions; operationalisations; review

Introduction

In the last few decades, engaging students in the thinking processes and activities of scientists – often referred to as scientific inquiry or inquiry-based instruction – has become a fundamental approach in science teaching and learning (National Research Council, 1996, 2012). Due to its importance, a huge body of research regarding the effectiveness of scientific inquiry exists – resulting, however, to some extent in inconclusive findings (for an overview, see e.g. Blanchard et al., 2010). Although considerable evidence exists that inquiry-based instruction positively affects different outcome measures including

CONTACT Silke Rönnebeck  roennebeck@ipn.uni-kiel.de

 Supplemental data for this article can be accessed here <http://dx.doi.org/10.1080/03057267.2016.1206351>.

© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

cognitive achievement, conceptual understanding, process skills, critical thinking and attitudes towards science (Anderson, 2002; Blanchard et al., 2010; Furtak, Seidel, Iverson, & Briggs, 2012; Haury, 1993; Minner, Levy, & Century, 2010; Schroeder, Scott, Tolson, Huang, & Lee, 2007), critics of inquiry-based teaching have repeatedly challenged its efficacy (Kirschner, Sweller, & Clark, 2006; Klahr & Nigam, 2004). Part of the disagreement may be due to the fact that the term inquiry has taken on different meanings within the science education literature. Inquiry refers not only to an instructional approach but also to curriculum materials, a way for students to learn science and scientific ways of obtaining knowledge (Bybee, 2000; Furtak et al., 2012). Moreover, even when focusing on inquiry as an instructional approach, considerable disagreement can be observed among researchers and educators with respect to its definition (Blanchard et al., 2010; Furtak, Shavelson, Shemwell, & Figueroa, 2012; Hmelo-Silver, Duncan, & Chinn, 2007), ranging from minimally guided, discovery-oriented approaches in which students engage in hands-on activities (e.g. Kirschner et al., 2006) to elaborate lists of actions for the students and their teachers (e.g. National Research Council, 1996). In recent years, the situation has become even more complicated since the field of science education in the United States has moved away from using the term inquiry and now refers to *scientific practices* (National Research Council, 2012). Thus, the terminology used to describe inquiry-based approaches in science teaching and learning is diverse. For the purpose of this review, these approaches are mainly subsumed under the term scientific inquiry to facilitate reading. However, when describing specific studies, the terminology employed in the study is used to reflect the diversity of the different approaches.

Looking at the conceptualisations of scientific inquiry found in the literature, two main dimensions of inquiry-based teaching can be distinguished: the type and range of activities that students engage in (e.g. Abd El Khalick et al., 2004; National Research Council, 2000) and the degree of guidance provided by the teacher (e.g. Furtak et al., 2012). Empirical studies investigating the effectiveness of scientific inquiry may vary considerably along these two dimensions. This is especially crucial for the validity of meta-analyses that attempt to synthesise the causal inferences made by individual studies. In a recent meta-analysis, Furtak et al. (2012) argue that 'insufficient attention has been given to the operationalization of the inquiry construct in the case of prior meta-analyses of inquiry-based teaching and that this has masked important differences in the efficacy of distinct features of this instructional approach' (p. 304). In their analysis, the authors introduced a framework for inquiry-based teaching that distinguished between cognitive features of the activity (i.e. procedural, epistemic, conceptual and social) and the degree of guidance given by the teacher. The cognitive features are described by specific activities that students conduct when they engage in scientific inquiry like, e.g. asking scientifically oriented questions (procedural), drawing conclusions based on evidence (epistemic) and arguing scientific ideas (social). Overall, the authors found a medium mean effect size; however, a considerable variability among effect sizes was observed when they were considered as a function of the cognitive and guidance dimensions of inquiry. The most positive effects were observed for activities related to the epistemic or a combination of the procedural, epistemic and social domains of inquiry; with respect to the guidance dimension, the results suggested that teacher-led inquiry lessons have a greater effect on student learning than those that are student led.

Earlier meta-analyses mostly relied on expansive definitions of inquiry-based teaching, often, however, without systematically addressing the differences in the conceptualisations of this instructional approach. Here, small to medium mean effect sizes were reported (Furtak et al., 2012). A similar medium mean effect was found in a meta-analysis by Schroeder et al. (2007) who defined inquiry teaching strategies quite generally as student-centred strategies requiring students to answer scientific research questions by analysing data.

Background and objectives of the review

In their meta-analysis of experimental and quasi-experimental studies of inquiry-based science teaching, Furtak et al. (2012) argued that 'coding inquiry as a dichotomy, as opposed to existing on a spectrum, fails to capture the range of activities and thinking processes in which students might be engaged'

(p. 304). They addressed this issue by categorising the studies according to the cognitive features of the inquiry activities. This implies, however, that these activities are defined and operationalised in a similar way across studies – which may not always be the case. In a study by Chen and Klar (1999), e.g. ‘students designed and evaluated experiments and made inferences from the experimental outcomes’ (p. 1098); McElhany and Linn (2011) similarly asked students ‘to design informative experiments’ and ‘to explain the mechanisms’ of a specific phenomenon (p. 746). Students in the former study, however, conducted hands-on experiments with real equipment while students in the latter study worked in a virtual experimentation environment. Following the argument by Furtak et al. (2012), as described above, it seems necessary to not focus solely on inquiry as a global concept, but on the different activities of the inquiry process in which students engage. The review thus aims to answer the question how coherently these activities have been defined and operationalised in empirical studies within the broader context of scientific inquiry.

In order to do so, a framework for inquiry-based teaching and learning in science is used that conceptualises scientific inquiry as a process consisting of activities that students conduct and the underlying competences that these activities require, respectively (e.g. Bell, Urhahne, Schanze, & Ploetzner, 2010; Linn, Davis, & Bell, 2004; National Research Council, 1996, 2000, 2012; Pedaste et al., 2015). Despite the widespread use of this approach in the science education literature, however, research varies considerably with respect to both the activities that are regarded as central to the process of scientific inquiry and especially the terminology that is used to label those activities (Abd El Khalick et al., 2004; Pedaste et al., 2015).

One of the most prominent lists of activities stems from the publications of the National Research Council. The framework for K-12 science education (National Research Council, 2012) lists eight scientific practices: 1. Asking questions, 2. Developing and using models, 3. Planning and carrying out investigations, 4. Analysing and interpreting data, 5. Using mathematics and computational thinking, 6. Constructing explanations, 7. Engaging in argument from evidence and 8. Obtaining, evaluating and communicating information. It is explicitly stressed that the eight practices are not separate but that they intentionally overlap and interconnect.

Other well-known models in the field distinguish specific phases in the inquiry process. Examples are, e.g. the 5E learning cycle model (Bybee et al., 2006) that lists five inquiry phases (engagement, exploration, explanation, elaboration and evaluation) and the inquiry cycle proposed by White and Frederiksen (1998) that also identifies five phases but labels them as question, predict, experiment, model and apply. In a study analysing models, tools and challenges of collaborative inquiry learning, Bell, Urhahne, Schanze and Ploetzner (2010) compared the specifications used in prominent models of inquiry learning with the aim of finding commonalities. They came up with nine categories of main inquiry activities (labelled as *processes*): orienting and asking questions, hypotheses generation, planning, investigation, analysis and interpretation, model, conclusion and evaluation, communication and prediction. In a recent review, Pedaste et al. (2015) tried to further systematise the various terminologies used to describe the activities of the inquiry process in order to develop a synthesised inquiry cycle. They distinguished five general *phases* of which three are divided into *sub-phases*: orientation, conceptualisation (divided into the sub-phases questioning and hypothesis generation), investigation (divided into the sub-phases exploration, experimentation and interpretation), conclusion and discussion (divided into the sub-phases reflection and communication). The latter two studies stress that the inquiry process does not imply a fixed chronological order of the different activities but that there are multiple possible pathways including sub-cycles and repetitions; moreover, in both models, communication is regarded as an overarching ability that is important for all steps of the process.

The model used in the analyses presented in this review is a synthesis of existing activity-based conceptualisations of scientific inquiry in the literature (Abd El Khalick et al., 2004; Bell et al., 2010; Bybee et al., 2006; Linn et al., 2004; Mullis, Martin, Ruddock, O’Sullivan, & Preuschoff, 2009; National Research Council, 1996, 2000, 2012; Pedaste et al., 2015; White & Frederiksen, 1998). It distinguishes nine activities: 1. Identifying questions, 2. Searching for information, 3. Formulating hypotheses and generating predictions, 4. Planning, designing and carrying out investigations, 5. Analysing, interpreting

and evaluating data, 6. Developing explanations, 7. Constructing models, 8. Engaging in argumentation and reasoning and 9. Communicating. In line with earlier models, no fixed chronological order is implied and the activities might overlap and interconnect.

Some of these inquiry activities have evolved as research fields of their own during the last decades, resulting in dedicated reviews. Reviews of research in modelling and argumentation provide some evidence that the variation observed in the definitions of scientific inquiry can also be found at the level of distinctive inquiry activities (Cavagnetto, 2010; Jiménez-Aleixandre & Erduran, 2007; Nicolaou & Constantinou, 2014). In a recent review of the assessment of modelling competence, Nicolaou and Constantinou (2014), e.g. found that studies usually address only parts of what can be conceptualised as modelling competence and often different definitions are used even when focusing on a common aspect. In the field of argumentation, Jiménez-Aleixandre and Erduran (2007) reviewed meanings of argument and argumentation in the literature. According to their review, different understandings of argument as well as argumentation exist. Whereas some authors agree that argument has both an individual (referring to any piece of reasoned discourse) and a social (referring to a dispute or debate between people) meaning, others restrict argument to the social meaning. With respect to scientific argumentation, again, two different viewpoints exist: argumentation as knowledge justification and argumentation as persuasion. Whereas the former is defined as the process of connecting claims and evidence through justification, the latter is related to the process of convincing an audience. This discussion is closely related to the question whether or not one should distinguish argumentation and explanation. Osborne and Patterson (2011) argue for the importance of this distinction because explanation and argumentation have different goals: 'Explanations and the construction of explanations are essential to the creation of new knowledge. The pedagogic value of argumentation, however, lies in its value for exploring the justification of belief and promoting a dialectic between construction and critique' (p. 636). Other authors, however, conflate the two constructs by mixing elements of arguments and explanations (e.g. McNeill & Krajcik, 2008). In his review of argument interventions in the context of scientific literacy, Cavagnetto (2010) found that the interventions varied with respect to the nature and purpose of the activity and the aspects of science included in it. The learning of argument is approached using three strategies: immersion in practice, explicit instruction in the structure of argument and emphasis of the interaction of science and society. In immersion-oriented interventions, argument constitutes an integral part of investigations; it is not considered as something that concludes an investigation but is present throughout the process as students identify questions, carry out experiments, interpret data and defend evidence-based knowledge claims. On the contrary, structure-oriented interventions focus on explanatory activities and separate argument and investigations; argument is thus considered more a product of investigations than an enmeshed component. Science-and-society-oriented interventions eventually use socio-scientific issues to contextualise argument (Cavagnetto, 2010). These results from the fields of argumentation and modelling show that the definitions and operationalisations of these inquiry activities vary considerably across studies. To our knowledge, no reviews for the other inquiry activities exist which address this issue of diversity in definitions, conceptions and/or operationalisations. The purpose of the present article is thus to take a first step towards closing this gap by providing insight into the range and spectrum of definitions and operationalisations related to activities of the inquiry process that have been used in empirical studies within the field of scientific inquiry.

Instead of looking at scientific inquiry from a holistic perspective, this review thus takes a rather atomistic approach (see Figure 1). It analyses the operationalisations of the different inquiry activities in empirical studies with respect to three major aspects: (1) How are the activities defined? (2) How are the activities implemented in the learning environment or intervention? and (3) How are students' competences with respect to these activities assessed? In other words, the analysis intends to extract from the reviewed empirical studies the operationalisations of the different inquiry activities as they become manifest in the theoretical considerations of the authors (i.e. their definitions), in the implementation of the activities in form of instructional or interventional components of the study and/or in the focus of an assessment of student performance related to the activities.

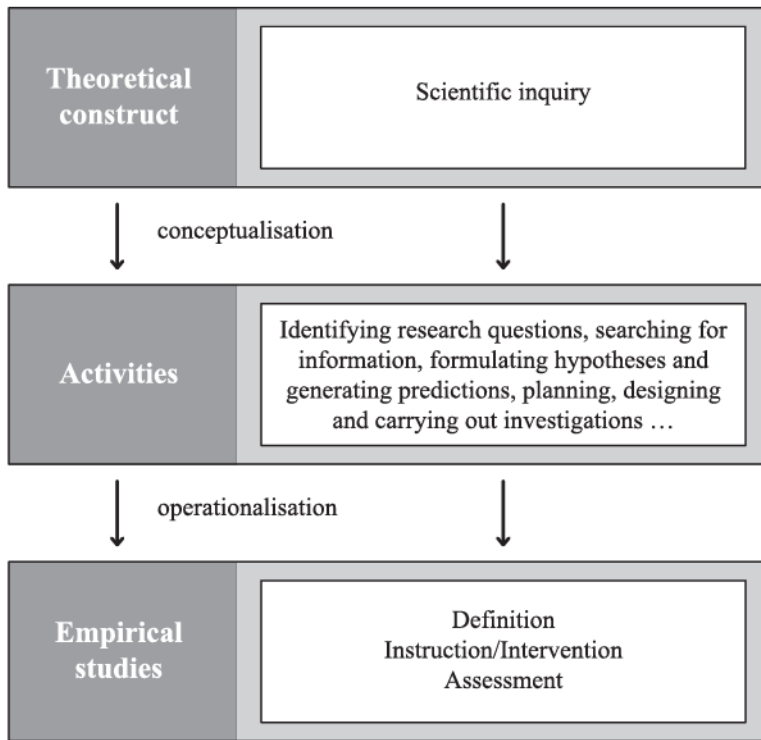


Figure 1. Analysing the construct scientific inquiry on different levels in this review.

Using such an atomistic approach is based on the assumption that understanding the different parts is a prerequisite for better understanding the whole, acknowledging, however, that the whole is certainly more than the sum of its parts. If we want to arrive at a more uniform, coherent and holistic understanding of scientific inquiry though, we first need to understand the different activities that are commonly considered to be the basis of this concept (Bell et al., 2010; Linn et al., 2004; National Research Council, 1996, 2000, 2012; Pedaste et al., 2015).

Method

One of the main challenges for a literature review is to ensure that as few as possible relevant publications are missed in the literature search. This is especially true if the research field is as diverse as in the case of scientific inquiry where a rich vocabulary exists to describe inquiry-related approaches in instruction, including, e.g. scientific inquiry, inquiry-based teaching and learning, authentic inquiry, project-based science, modelling and argumentation, hands-on science and constructivist science (Furtak et al., 2012). One strategy to address this challenge is to apply search criteria that are as broad and comprehensive as possible. The underlying idea is to generate an initial literature database that in case of doubt includes publications that are not related to the objectives of the review rather than risking missing important contributions to the field of interest. At the same time, however, the resulting initial number of publications must still be reasonable and manageable for further analysis with regard to inclusion and exclusion criteria. In the case of this review, different search strategies were pursued – namely searching in relevant databases, in relevant journals and in the reference lists of relevant publications found using the first two strategies – and by carefully selecting the keywords to be used in the database searches.

For the choice of keywords, we first analysed previous reviews in the field of scientific inquiry with respect to the keywords they had used (e.g. Furtak et al., 2012; Heinz, Lipowsky, Gröschner, & Seidel, 2012). Second, trial searches were conducted with different combinations of these keywords that varied with respect to their spectrum, i.e. the extent to which they included alternative teaching and learning approaches that are related to scientific inquiry. The aim of these trials was to find a reasonable compromise between comprehensiveness, on the one hand, and size of the initial literature database for further analysis, on the other. Two library databases, Web of Science and ERIC, were searched to provide this initial database. The search was restricted to publications that (a) were published before 1 April 2013 and after 1 April 1998 and (b) were written in English. The following keywords were finally chosen for describing the approach of scientific inquiry: inquiry, collaborative learning, discovery learning, cooperative learning, constructivist teaching, problem-based learning and argumentation. Since the focus of this review is on empirical research in scientific inquiry from K-12, these keywords were crossed with the following keywords representing the area of evaluation and assessment: assessment, evaluation, validation, achievement or feedback and discourse, effective questioning, assessment conversations, accountable talk, quizzes, self-assessment, peer assessment, portfolio, learn log, mind map, concept map, rubrics, science notebook, multiple-choice, constructed response or open-ended response. Additional keywords were used to further limit the selection to the subject of science at the school level. After eliminating duplicates, the search led to a sample of $N = 331$ publications (see Figure 2). In a second step, those journals that appeared to produce the greatest number of articles in the database searches (Journal of Research in Science Teaching and Science Education) were specifically examined to ensure that all the relevant literature they offered is included in the literature database for this review. Since the review has a special focus on empirical research, three journals from the field of educational assessment (Applied Measurement in Education, Assessment in Education and Educational Assessment) were also included in this second step. Eventually, the reference lists of selected articles were searched for relevant articles not already in the database.

Combining these three steps led to an initial literature database of $N = 459$ publications. These studies were then further analysed by reading abstracts and, if necessary, full texts with respect to the following inclusion criteria: (a) studies are based on empirical data, (b) are related to scientific inquiry, (c) are situated at the school level, i.e. in kindergarten, primary, lower or upper secondary level, and (d) were published in a peer-reviewed journal. There were several reasons for the focus on journal articles, the most important being the peer review process since this provides some check of quality of the presented research. Moreover, journal articles are the type of literature best accessible using systematic searching procedures. Applying these criteria led to a final database of $N = 96$ publications (the complete search and selection process is depicted in Figure 2).

These $N = 96$ publications were then read by the authors and analysed with respect to general information about the studies (e.g. year of publication, country, type of study and sample characteristics) and the activities of the inquiry process that they addressed either as part of the learning environment (e.g. curriculum or instructional unit) or as part of the assessment employed in the study. Most studies in the review emphasise particular inquiry activities. However, 15 studies have a more general focus on the effects of implementing inquiry in K-12 classrooms, in particular the impact of different inquiry-oriented curricula or instructional approaches. The output measures in these studies differ between students' content knowledge and conceptual understanding, views on the nature of science, attitudes, interest and motivation. Although these studies are relevant for this review from the perspective that they are integrated in an inquiry-oriented theoretical background, they often provide only few details about the implementation of inquiry activities in their designs. The information about students' activities is often rather general like, e.g. 'students use inquiry-based curricula and an internet software programme to study general weather topics such as wind, precipitation, temperature and pressure, and clouds and humidity collaboratively with students and professional scientists' (Mistler Jackson & Songer, 2000, p. 464). This is not to criticise these studies or publications, but details about the implementation of scientific inquiry in the classroom or learning environment are a condition precedent to this review. Hence, these 15 studies were excluded from the subsequent presentation of results. The final database

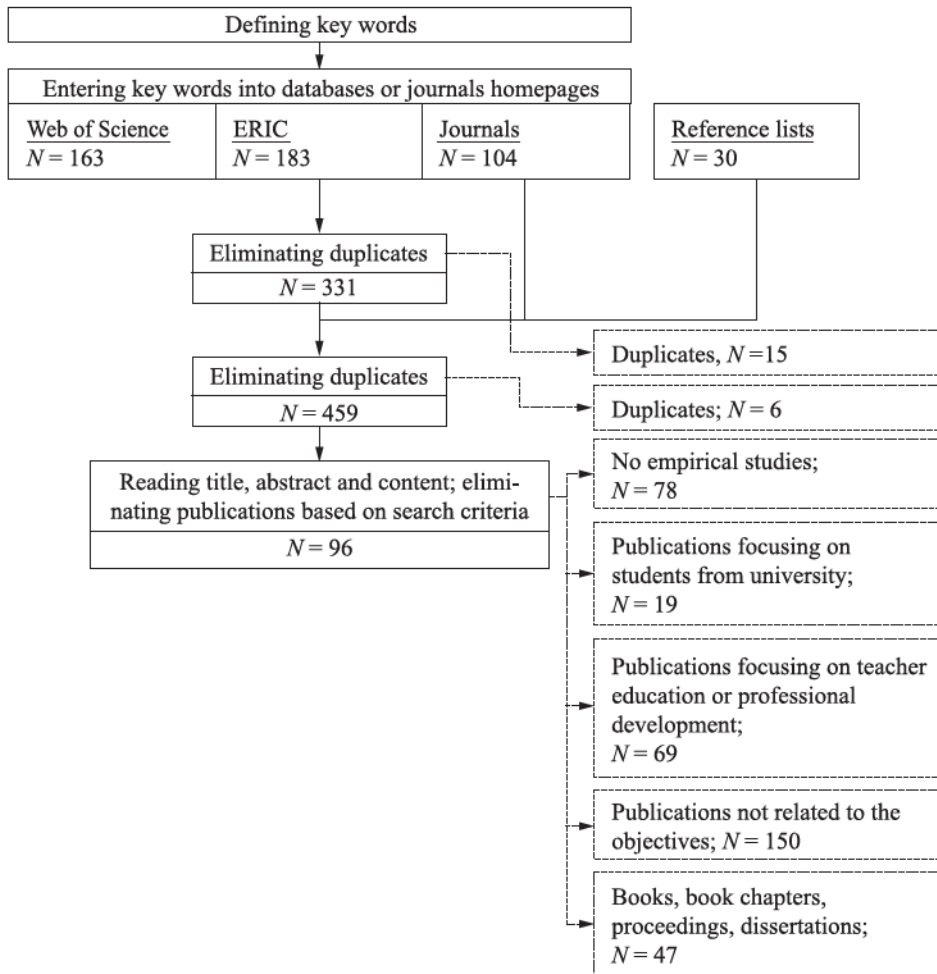


Figure 2. Schematic representation of the literature search and selection process.

thus consisted of $N = 81$ publications. A complete list of all publications analysed in this review with descriptive information about the studies can be found in the online supplementary material.

Analysis

This review intends to provide insights into the range and spectrum of definitions and operationalisations used with regard to single activities of the inquiry process in the different studies. The results of the analysis are presented in nine sections related to nine inquiry activities, acknowledging, however, that there is not always a sharp distinction and that in practice, these activities are often very closely related (National Research Council, 2013). In each section, information is provided with respect to the theoretical background as well as the operationalisation – both with respect to the implementation of the activity in the learning environment and the assessment of students' competences related to a specific activity.

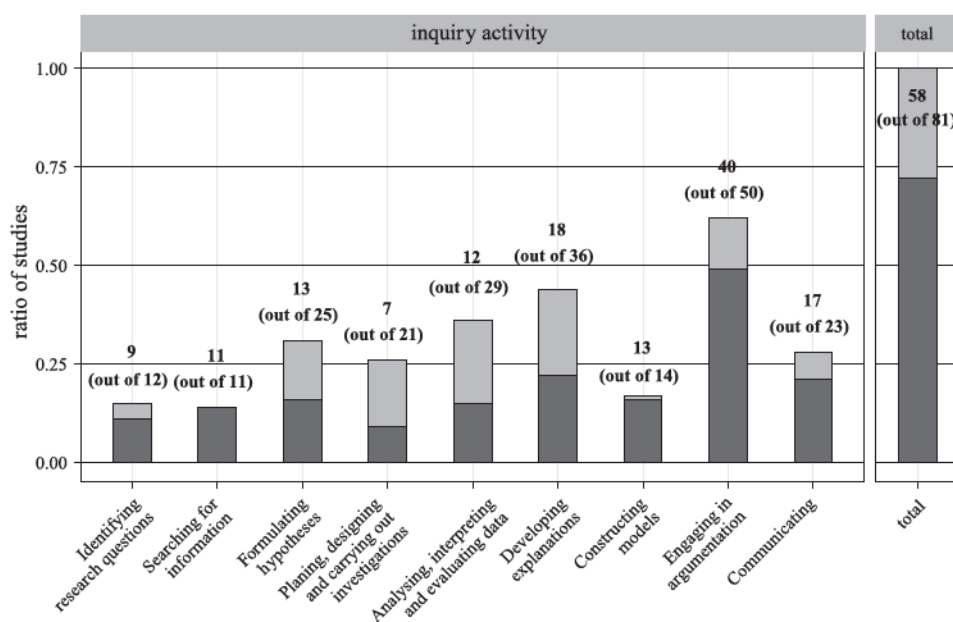


Figure 3. Distribution of studies focusing on specific inquiry activities and providing conceptual details of this activity. The larger bars (light grey) indicate the ratio of studies addressing a specific activity in relation to the total number of studies reviewed ($n = 81$). The dark grey bars indicate the portion of studies providing conceptual details within each activity category. The annotation above each pair of bars provides the absolute numbers for this comparison, giving the number of studies providing conceptual details and the total number of studies for each activity category, e.g. 12 studies focused on the inquiry activity of identifying questions and 9 out of these 12 studies provided conceptual details about this activity.

Identifying research questions

For the analyses within this review, the activity of identifying research questions was distinguished from the more general and content- (or comprehension-) related activities of questioning (e.g. Chin & Osborne, 2010) and question-posing (e.g. Kaberman & Dori, 2009). Questioning requires students to 'engage with their current understanding, probe into alternative ways of explaining phenomena, and ask why certain explanations are better than others' (Chin & Osborne, 2010, p. 886). Aguiar, Mortimer and Scott (2010) call this kind of questions *wonderment questions* because they 'require integration of complex and divergent information from various sources, and reflect curiosity, puzzlement, scepticism or speculation' (p. 175).

The focus of this review, however, is on the identification of research questions. Twelve publications address this issue, but details concerning this construct are provided in only nine of them (see Figure 3). An explicit definition of a research question is only given in the study by White and Frederiksen (1998) where students should formulate 'a well-formed, investigable research question whose pursuit will advance their understanding of a topic they are curious about' (p. 10). Other studies mainly focus on one of the two characteristics, respectively, the need for research questions to be testable (Chang et al., 2011; Ebenezer, Kaya, & Ebenezer, 2011) or their potential to advance understanding (Cavagnetto, Hand, & Norton-Meier, 2010). A third characteristic is eventually addressed in a study by Samarapungavan, Patrick and Mantzicopoulos (2011) who focus on students' ability to use science concepts in the generation of scientific research questions.

A specific focus on the identification of research questions is found in a study by Hofstein, Navon, Kipnis and Mamlok-Naaman (2005). They investigated the effects of inquiry-type laboratory activities on students' ability to ask more and better questions and to choose research questions for further research. The results showed that students improved their ability to ask better and more relevant questions as a result of gaining experience with the inquiry-type experiments. Additional studies assessed the ability

to identify research questions as part of students' inquiry abilities, mostly with respect to interventions fostering these abilities. Ebenezer et al. (2011), e.g. analysed the effects of participating in long-term scientific research projects on inquiry abilities using rubrics consisting of 11 criteria to assess students' project reports. Their analyses showed that students reached the highest proficiency values with respect to those two criteria that were related to the formulation of research questions. In a similar way, So (2003) included students' ability to judge primary students' research reports in a survey study. In two studies by Samarapungavan, Mantzicopoulos, and Patrick (2008) and Samarapungavan et al. (2011), dealing with the learning of science through inquiry in kindergarten, an electronic portfolio system was used to collect and evaluate evidence of children's learning through classroom inquiry activities. The portfolios contained two types of data, student artefacts (e.g. records in science notebooks or posters) and digital videos and transcriptions of the intervention activities. One criterion for the portfolio analysis was the raising of research questions and predictions (Samarapungavan et al., 2011). Results showed that the inquiry intervention led to significant improvements with respect to children's ability to raise research questions and predictions. Finally, students' ability to formulate research questions was included as one facet of inquiry competence in a self-report questionnaire that was evaluated in a study by Chang et al. (2011).

In summary, all studies focus on students' ability to identify or raise research questions, mainly by means of open-ended formats or portfolios. To evaluate the quality of these questions, different characteristics are used, e.g. the need for research questions to be testable (Chang et al., 2011; Ebenezer et al., 2011), their potential to advance understanding (Cavagnetto et al., 2010) or the application of science concepts in the generation of research questions (Samarapungavan et al., 2011). However, no study focuses on the question how the assessment format impacts the evaluation of students' abilities to identify and raise research questions and whether there is a preferable format when focusing on specific aspects of this activity. Few studies address fostering students' abilities in identifying research questions. Moreover, these studies focus entirely on an immersion approach of participating in inquiry-type laboratory activities or research-like projects. Hence, apart from repeated practice, little is known about instructional activities to develop students' ability to identify research questions, in general, as well as with regard to the different characteristics of this inquiry activity.

Searching for information

A focus on searching for information is often related to ill-structured, real-life problems in collaborative learning environments like, e.g. project-based science (Butler & Lumpe, 2008; So, 2003), problem-based learning or problem-solving activities in general (e.g. Belland, Glazewski, & Richardson, 2011; Chiou, Hwang, & Tseng, 2009; Simons & Klein, 2007; Toth, Suthers, & Lesgold, 2002; Tsai, Hwang, Tsai, Hung, & Huang, 2012; Wong & Day, 2009) and computer-based virtual collaborative environments (e.g. Ketelhut & Nelson, 2010; Taasooobshirazi, Zuiker, Anderson, & Hickey, 2006). Eleven studies in this review investigate and provide further details about this inquiry activity (see Figure 3). Searching for information is mostly addressed as one activity – among others – that contributes to a problem-solving or inquiry process. Students are required to search different sources for information that may help them solve the problem at hand (e.g. Taasooobshirazi et al., 2006). Common sources are digital libraries (Butler & Lumpe, 2008; Tsai et al., 2012), links to relevant websites (Simons & Klein, 2007), school libraries (Wong & Day, 2009) and interactions with virtual characters like, e.g. computerised residents in multi-user virtual environments (Ketelhut & Nelson, 2010; Spires, Rowe, Mott, & Lester, 2011).

There are, however, also studies that place specific emphasis on the search process and the underlying strategies. Here, two lines of research can be identified. The first line is related to scaffolding which should help students determine what information is needed, how to find this information and how to organise it (Belland et al., 2011; Butler & Lumpe, 2008; Simons & Klein, 2007). Butler and Lumpe (2008), e.g. analysed the use and effects of computer scaffolds. In a project-based science unit, searching features (i.e. how often do students perform a search, use the dictionary, use the thesaurus, read the website description or view the actual website) were investigated as one of the five scaffolding

categories to result in a descriptive statistic of scaffold use. The authors found that students used the searching features more than half of the time and more often than any other scaffolding category. Moreover, a significant correlation between the use of the searching features and the student scores for self-efficacy for learning and performance was observed.

The second line of research is related to the computer-assisted analysis of students' search behaviour and their underlying problem-solving abilities (Chiou et al., 2009; Toth et al., 2002; Tsai et al., 2012). Toth et al. (2002), e.g. focused on the selection and evaluation of evidence from multiple sources. Students had to search hypertext-based, simplified research papers for hypotheses and data and establish links between them. The resulting information search measure was based on the number of topic-relevant pieces of information that had been recorded and how many of these had been labelled as data and hypotheses. The study compared two technology-based knowledge representation tools, evidence mapping and prose writing. The authors found that the mean number of labelled information pieces was significantly higher in the mapping groups than in the prose groups. The difference between treatment conditions was attributed not to students' ability to categorise information into hypotheses and data, but to their explicit recognition of the necessity to do so. An automatic scoring mechanism to assist teachers in evaluating the web-based information-searching and problem-solving ability of individual students was developed and evaluated by Chiou et al. (2009). Their analysis of students' information-searching behaviour was based on the Big6 model: task definition, information-seeking strategies, location and access, use of information, synthesis and evaluation. A correlation analysis resulted in large positive correlations between teacher and automatic scores for all indicators except information-seeking strategies.

In essence, the different studies focus either on the information or the search aspect of this inquiry activity. When focusing on the information aspect, students' ability is often evaluated with regard to the degree to which the collected information contributes to the problem-solving or inquiry process. Other studies are interested either in investigating students' search behaviour (e.g. by log files of computer-based learning environments) or in identifying means to scaffold and support students' search process (e.g. by providing strategies to select, process and organise the contextually relevant information). Both lines of research often make use of ill-structured problems, collaborative learning environments and multiple resources (digital or traditional libraries, web quests, etc.) and focus mainly on describing the students' search behaviour, while only little emphasis can be found with regard to the assessment of this activity (cf. Toth et al., 2002).

Formulating hypotheses and generating predictions

In total, students' ability to formulate hypotheses or generate predictions is explicitly addressed in 25 publications. Despite this large number of studies, only 13 studies disentangle this aspect of inquiry in detail (see Figure 3). In the other 12 studies, the formulation of hypotheses is mentioned as an important aspect of inquiry, but little detail is given about its function and operationalisation in the learning environment or the assessment.

Regarding the definition, hypotheses are seen as the relation between input and output variables (Gijlers & de Jong, 2005). The main purpose of formulating hypotheses is often stated as to allow students 'to learn and experience science with greater understanding and to practice their metacognitive abilities' and to provide them 'with the opportunity to construct their knowledge by actually doing scientific work' (Hofstein et al., 2005, p. 795). However, within the reviewed studies, students' perspective on the function of generating hypotheses is seldom addressed. Herrenkohl, Palincsar, DeWater and Kawasaki (1999) asked students about the function of formulating this kind of predictions, but the coding of students' answers was limited to decide whether these answers were at least at the level of *guess* or *educated guess*.

Different formats are used to assess students' ability to formulate hypotheses or generate predictions. These range from closed multiple-choice formats (Gobert, Pallant, & Daniels, 2010) to open-ended items (Furtak & Ruiz-Primo, 2008), students' discourse (Gijlers & de Jong, 2005) and are part of conducting

hands-on (Hofstein et al., 2005), computer-based (Ketelhut & Nelson, 2010) or thought experiments (Herrenkohl, Tasker, & White, 2011).

Studies varied according to the evaluation of the quality of students' hypotheses. If details were provided, most studies differentiated between hypotheses that are testable (i.e. *correct* hypotheses) and those that are not. With regard to students' ability in formulating a testable hypothesis, Ebenezer et al. (2011) expect students to 'be able to state a hypothesis that lends itself to testing. Also, the hypothesis should be accompanied by coherent explanation(s)' (p. 103). A detailed taxonomy is provided by Kaberman and Dori (2009) who differentiated content (whether only the phenomenon at hand or a more general level was addressed), thinking level (according to Bloom's taxonomy) and chemistry understanding levels (macroscopic, microscopic, symbolic and process levels). Findings suggest that both the number and the complexity of students' hypotheses increased due to an intervention based on this framework (Kaberman & Dori, 2009).

Several interventions have been suggested to promote students' ability in formulating hypotheses. Spires et al. (2011) used a gameplay approach that required solving a science mystery based on microbiology content: 'Results indicated that the effective exploration and navigation of the hypothesis space [...] was predictive of student learning' (Spires et al., 2011, p. 453). Using constructed response items, Lavoie (1999) examined the effects of adding a prediction or discussion phase where students individually wrote out predictions with explanatory hypotheses at the beginning of a learning cycle. By introducing this phase, the author intended to prompt students to construct and deconstruct their procedural and declarative knowledge. The evaluation of this intervention revealed significant gains with respect to process and logical thinking skills, the understanding of scientific concepts and students' attitudes towards science.

Kyza (2009) examined students' inquiry practices in considering alternative hypotheses. She analysed students' discourse, actions, inquiry products and interactions with their teachers and peers. Despite significant learning gains when implementing a supportive learning environment (i.e. teacher- and task-based scaffolding), the author pointed out several epistemological problems related to students' perception of the usefulness of examining and communicating alternative explanations, i.e. by relying primarily on a verification strategy of hypothesis testing. Her findings indicate the importance of epistemologically targeted discourse alongside guided inquiry experiences for overcoming these challenges.

Throughout the reviewed studies, formulating hypotheses is regarded as a core feature of scientific inquiry and as highly important to learn and experience science with greater understanding (cf. Hofstein et al., 2005). As mentioned above, however, few details about the function and operationalisation of formulating hypotheses in the learning process are provided. In addition, students' perspectives on the function of generating hypotheses and the influence of their perception on the whole inquiry process are barely addressed. Kyza (2009) pointed out that students tend to rely primarily on a verification strategy of hypothesis testing, indicating epistemological constraints in students' perception and interpretation of the role of hypotheses (and also alternative hypotheses) in the inquiry process.

Across the studies, a large range of different formats is used to assess students' abilities to formulate hypotheses (e.g. multiple choice, students' discourse or thought experiments). However, in most cases, the evaluation is restricted to the decision whether the proposed hypothesis is testable or not. Likewise, approaches to promote students' ability in formulating hypotheses are predominantly based on repeated practice while more detailed and focused instructional approaches (e.g. Kaberman & Dori, 2009) are hard to find.

Planning, designing and carrying out investigations

In total, 21 publications addressed the activity of planning, designing and carrying out investigations and again only the minority of papers ($n = 7$) pointed out details about what was expected from students regarding this scientific practice (see Figure 3). According to Ebenezer et al. (2011), designing and conducting scientific investigations means 'that students should logically outline methods and procedures, use proper measuring equipment, heed safety precautions, and conduct a sufficient number of

repeated trials to validate the results' (p. 103). Several publications investigated the activity of designing an investigation; however, in most cases, students' approaches were limited by predefined guidelines. Few studies were found in which students were unrestricted in deciding about the scope and design of their investigation. Thus, instructional decisions concerning the implementation of this activity seem to almost automatically entail aspects of scaffolding and guidance.

Chen and Klahr (1999) predominantly focused on the control of variables strategy and how students can be supported to generalise this processing strategy across various contexts. They asked children in primary school to design and evaluate experiments and to make inferences from the experiment outcomes:

When provided with explicit training within domains, combined with probe questions, children were able to learn and transfer the basic strategy for designing unconfounded experiments. Providing probes without direct instruction, however, did not improve children's ability to design unconfounded experiments and make valid inferences. (Chen & Klahr, 1999, p. 1098)

According to the authors, the ability to transfer learned strategies to remote situations seems to increase with age.

Two other activities were present in the different publications next to designing an investigation: either students were asked to manipulate variables in a given experimental set-up (e.g. in a computer-based simulation environment; Valanides & Angeli, 2008) or they were asked to interpret an investigation designed by others. For instance, Zion, Michalsky and Mevarech (2005) confronted students with a phenomenon, findings collected by scientists that described the phenomenon and the experiments designed by scientists for solving the problem: 'Students were required to identify the relevant variables, interpret the results of the given experiment and draw valid conclusions on the basis of the given data' (Zion et al., 2005, p. 967).

In addition to the activities students were asked to perform, the reviewed publications also differed according to the mode in which students realised the planning of their investigations. Three major specifications could be identified: hands-on, virtual and theoretical. Hands-on experiments account for the majority of publications (e.g. Chen & Klahr, 1999; Dori, 2003). In most cases, students were provided with technical equipment and are responsible for designing, setting up and conducting the experiment. Other studies used surrogates to the technical, hands-on realisation of scientific experiments by using computer-based systems. For example, McElhaney and Linn (2011) developed a computer simulation in which students conducted experiments to answer different questions. The questions could be selected from a drop down menu or students could choose an alternative such as *just exploring*. While students conducted their experiments, the software logged the question and the variable values that the students selected for each trial. The question students chose was used to infer their aims in each trial. The third group of publications included a theoretical approach to designing an experiment, i.e. students were asked to outline an experiment in written form, mainly as part of assessing students' inquiry abilities. For instance, Yoon (2009) used the Diet Cola Test which requires students to specify a research question related to a given situation and to design an experiment to find the answer. In this approach, students' ability to design an experiment is often treated as an isolated step, i.e. subsequent steps of data analysis and interpretation are unrelated to the students' experimental design in this specific step.

Irrespective of the mode of investigation, students were confronted with different degrees of openness in the different studies or, as mentioned above, with different levels of scaffolding and guidance. In the case of hands-on experiments, often, the kind and amount of technical apparatuses were pre-selected either to guide the students or to prevent danger. Also some virtual set-ups allowed students 'not only [to] design the hypothesis, but also the procedure and data-collection methodology' (Ketelhut & Nelson, 2010), while other systems were more restricted so that students' design of experiments was limited to, e.g. manipulating 'values of input variables, and [observing] the behaviour of output variables' (Gijlers & de Jong, 2005). Regarding differences between the different modes and degrees of openness, Stecher et al. (2000) investigated whether the content domain, the format (paper and pencil vs. hands-on) and the level of inquiry (whether the task guided the student or required the development of a solution strategy) had an impact on students' performance. The authors used a shell design

to develop six similar investigations of acids, controlling for format and level of inquiry. However, '*post hoc* analyses of the tasks revealed unanticipated differences in developers' interpretation of the shell that may have affected student performance' (Stecher et al., 2000, p. 140). The authors concluded that comparing students' performance across the different modes and levels of inquiry seems more difficult than expected as students' performance within each mode and level also varies to a large extent.

Only two publications were identified that tried to assess students' ability in designing an experiment via a multiple-choice test. Gijlers and de Jong (2005) used ten multiple-choice items to examine students' performance in the areas of planning and conducting an investigation. Items in this section of the test aimed at the identification of relevant variables, the design of an experiment, the ability to state a hypothesis, and identification of data that support a hypothesis. (p. 271)

Chang et al. (2011) focused on students' self-evaluation of their ability to design and conduct experiments. The authors asked students whether they considered themselves able to 'adopt a suitable strategy for specific questions or hypotheses, employ resources, and then work out a problem-solving approach' (Chang et al., 2011, p. 1220). The authors reported high latent correlations to students' self-reported abilities of formulating a hypothesis and analysing data but only a medium correlation to conducting an experiment (Chang et al., 2011).

Regarding possibilities to foster students' abilities in designing and conducting experiments, White and Frederiksen (1998) investigated the effect of reflective assessment on inquiry units. Overall students' performance improved significantly and a controlled comparison revealed that students' learning was greatly facilitated by reflective assessment. Interestingly, adding this metacognitive process to the curriculum was particularly beneficial for low-achieving students: performance in their research projects and inquiry tests was significantly closer to that of high-achieving students than was the case in the control classes.

As in the case of studies on students' ability to formulate hypotheses, only the minority of studies reviewed with regard to investigating students' planning, designing and carrying out of experiments provide details about the expected outcome. Three main lines are identified in these investigations: students are asked to design an investigation, to manipulate a given set-up or to interpret a set-up designed by others. Within these approaches, different modes are used in which students' ability is assessed: hands-on, virtual and theoretical. While designing an experiment is assessed in the full range of different modes, manipulating a given set-up is mostly investigated in virtual environments and interpreting a set-up designed by others mainly in written form. In summary, the impact of both the approach and the mode on the obtained results remains difficult to evaluate. Consequently, the degree to which results obtained in different settings are comparable remains unclear (cf. Stecher et al., 2000).

With regard to fostering students' ability in this inquiry activity, most studies seem to automatically entail aspects of scaffolding and guidance, e.g. by providing students with predefined guidelines or preselected designs and materials. The degrees of openness and scaffolding, however, vary widely.

Analysing, interpreting and evaluating data

The evaluation of results is included in many publications as a step of inquiry but often only as a buzzword or by-product of a more general view on inquiry. Hence, few studies aim to describe the steps that must be taken to collect data that can be interpreted in a scientific way. Among the studies included in this review, while 29 studies address this scientific practice, only 12 publications provide details.

According to Chang et al. (2011), students should analyse data and establish evidence, build the link between evidence and conclusion and then establish the relationship between evidence and conclusion to form a model or explanation through logical thinking. For these steps, appropriate tools, methods and procedures are necessary to collect and analyse data systematically, accurately and rigorously. In some cases, this can include the use of mathematical tools and statistical software, e.g. to analyse and display data in charts or graphs or to test relationships between variables (Ebenezer et al., 2011).

Studies on students' ability to analyse and interpret data differ according to the activity students have to carry out (conduct their own analysis; evaluate a given analysis or interpretation; and/or self-evaluation of one's ability) and the mode of realisation (hands-on; virtual; and theoretical). Across all publications, students are predominantly required to conduct an own, hands-on analysis of self-collected data. For instance, Chen and Klahr (1999) asked students to make systematic comparisons to determine the effects of different variables on a spring. In each task, participants were asked to focus on a single outcome that was affected by four different variables. For example, the outcome was how far the spring stretched as a function of its length, width, wire size and weight (Chen & Klahr, 1999).

In a study conducted by Vellom and Anderson (1999), students learned about mass, volume and density by attempting to stack three miscible solutions with differing densities on top of one another. After several attempts, students were asked to decide in small groups which of the different claims students in the class made were trustworthy and which were unreliable, i.e. to decide how to 'separate the data from the noise' (Vellom & Anderson, 1999, p. 182). In terms of standards for assuring the quality of the collected data, replicability, care in experimentation, explicitness about experimental procedures and consistency of observed and reported results were pointed out by the students (Vellom & Anderson, 1999).

Analysing primary school children's inquiry approaches, So (2003) recognised that these

children often used daily commodities to measure or collect data, and used other equipment and instruments when needed, [...] children were able to make sense of their data by using scientific equipment and empirical observation, and to translate these observations into useable data for interpretation, as well as gathering data in an organized and logical manner, [...] and it was common to find from children's reports that they were capable of comparing the several rounds of data collected and to come to an agreement about the set of data for interpretation. (pp. 187–188)

Toth et al. (2002) identified patterns in students' inquiry approaches resembling two different strategies with respect to scientific reasoning. Some students followed a *reasoning from hypothesis* approach, while others started with collecting data following a *reasoning from data* approach to scientific reasoning (Toth et al., 2002). The authors concluded that different scaffolds may be needed to support students who tend to apply either one of these two approaches.

Several studies in this review used virtual, computer-based systems in their investigations. In the context of plate tectonics, Gobert et al. (2010) asked students to create cross sections of the earth's interior at different plate boundaries to elaborate on the magnitude, depth, frequency and location of earthquakes and to explain how the movements of the plates at each boundary account for patterns in each set of earthquake data. Students were also asked to apply their understanding to the reverse case, i.e. they were given two tables of earthquake data and were asked to identify the type of boundary represented by each table (Gobert et al., 2010).

In their study, Toth et al. (2002) used a *design experiment* approach to develop an instructional framework that lends itself to authentic scientific inquiry. A technology-based knowledge representation tool enabled students to relate hypotheses to data by constructing so-called evidence maps. Students formulated scientific statements using different shapes for *hypotheses* and *data* and indicated the relation between these with *for* (support) and *against* (refutation) links. Additionally *and* links could be used to conjoin statements. With regard to the evaluation of data in relation to theories, students using the evidence map outperformed their counterparts who used prose writing. This effect was even enhanced by the use of reflective assessment throughout the inquiry process (Toth et al., 2002).

Comparable to the planning and designing of experiments, students' ability to analyse and interpret data is analysed based on different activities students have to carry out (conduct an own analysis; evaluate a given analysis or interpretation; and/or self-evaluate one's ability) and different modes of realisation (hands-on; virtual; and theoretical). Across all publications, students are predominantly required to conduct hands-on analyses of self-collected data while the evaluation of a given analysis or the self-evaluation of one's ability is mainly assessed in computer-based or written form. However, again, few studies provide details about the steps required to collect data that can be interpreted in a scientific way. Regarding the evaluation of students' ability to analyse and interpret data, students' controlling of variables and the systematics of comparisons between cases are the main features. On a more epistemological level, Vellom

and Anderson (1999) asked students about standards to assure the quality of the data collection and analysis. These aspects of standards and good scientific practice are not addressed in any other study included in this review.

With the aim to foster students' ability to analyse and interpret data, most approaches focus on means to support students in incorporating contextually relevant theories or students' hypotheses into the process of analysing and interpreting the data, i.e. in *linking back* the analysis to previous steps in the inquiry process. Here, the use of evidence maps and reflective assessment has proven fruitful (Toth et al., 2002). However, students in the different studies are rarely confronted with conflicting evidence or with complex methods in the data analysis, indicating that the *outcome space* of experiments or provided data-sets is mainly controlled to focus on clean, clear-cut and well-structured results.

Developing explanations

The construction of evidence-based explanations is addressed in 36 publications in this review – however, only 18 of them further elaborate on the activity. Approximately half of these 18 studies address the development of explanations in the general context of scientific argumentation, emphasising the close relationship between explanation and argumentation (McNeill, 2009). The most detailed definition of a scientific explanation is given by Gotwals and Songer (2010):

We define a scientific explanation as a response to a scientific question that takes the form of a rhetorical argument and consists of three main parts: a claim (a statement that establishes the proposed answer to the question), evidence (data or observations that support the claim), and reasoning (the scientific principle that links the data to the claim and makes the reason visible why the evidence supports the claim). In short, a scientific explanation is a compilation of evidence elicited through observation and investigation and the explicit links those data have to related scientific knowledge. (p. 263)

This definition is closely related to the argumentation model by Toulmin (1958). References to this model can also be found in Cavagnetto et al. (2010) and Sampson, Grooms and Walker (2011). Whereas the former consider explanations as part of rebuttals, the latter regard them as one form of a claim (next to conclusions, conjectures or other answers to research questions). More content-oriented definitions understand explanations as a reference 'to how or why something happens' (McNeill, 2009, p. 235), as a form of schematic knowledge and kinds of mental models (Furtak & Ruiz-Primo, 2008) or as one aspect of constructing understanding (Wilson, Taylor, Kowalski, & Carlson, 2010). In the vast majority of publications, however, no explicit definition of an explanation is given and it is often not clearly separated from related activities like, e.g. drawing conclusions (Gobert et al., 2010).

Next to argumentation, developing explanations is also related to the construction and use of models. In general, models are regarded as support structures that allow students to develop explanations for phenomena, either by activating schematic knowledge (Furtak & Ruiz-Primo, 2008) or by providing 'a set of representations, rules, and reasoning structures' (Schwarz & White, 2005, p. 166). This relation is used by Wilson et al. (2010) who measured students' ability to reason with scientific models through constructed response items in which students were asked to explain or predict patterns in novel situations. Other studies consider explanations and models as alternative explanatory structures for phenomena, both of which are based on evidence (Ebenezer et al., 2011; Sampson et al., 2011).

The publications addressing students' scientific explanations differ with respect to the focus and the goal of their analyses. Whereas some studies explicitly focus on scientific explanations (either related to argumentation or not), others address explanations within the broader framework of argumentation and/or as one (of several) inquiry skills. Studies with an explicit focus on explanations clearly separate the content and structure of explanations in their analyses. Sampson et al. (2011) investigated the effect of an instructional model that requires students to develop, refine, evaluate and use explanations on students' argumentation and explanation. The effect was evaluated using a performance task that asked students to generate an original and complex written explanation (called argument) for an ill-defined problem. The task is coded according to four criteria: the adequacy of the explanation (regardless of its accuracy), the conceptual quality of the explanation, the quality of the evidence and the sufficiency

of the reasoning. Overall, the results indicate that the intervention increased students' disciplinary engagement and supported them in producing better arguments. Separate codes for content and structure can also be found in Gotwals and Songer (2010) and McNeill (2009), whereas Berland and Reiser (2009) focus solely on explanation structure.

Within the context of scientific argumentation, the publications mainly focus on the structure of explanations. Students' explanations are analysed applying Toulmin's model of argumentation as parts of rebuttals (Cavagnetto et al., 2010) or as part of a combined category consisting of Toulmin's data, warrants and backings called grounds (Clark & Sampson, 2008). Results indicate that students' rebuttals are often not fully developed rebuttals but rather objections to ideas (Cavagnetto et al., 2010). The analyses by Clark and Sampson (2008) moreover showed that students included grounds in their comments only half of the time. If they included some type of grounds, they mostly relied on an explanation without evidence and even if they included evidence, they mostly relied on simple justifications instead of coordinating multiple pieces of evidence. Other studies in this field focus, e.g. on students' ability to make event–evidence–explanation connections (Ebenezer et al., 2011) or on the nature of students' scientific thinking, i.e. 'how they reason, how they try to make sense of scientific ideas, and how they explain and justify answers that they give' (Steinberg, Cormier, & Fernandez, 2009, p. 020104-1).

Next to discourse analyses, another major approach to assessing students' explanations is based on the analysis of students' written responses. They are analysed based on different types of explanation items. Students are either explicitly asked to write their 'best explanation' for a specific phenomenon (Herrenkohl et al., 1999, p. 460), to provide explanations for their answers to multiple-choice items (Steinberg et al., 2009; White & Frederiksen, 1998) or to answer to assessment prompts, e.g. in the form of predict–observe–explain or constructed response items (Furtak & Ruiz-Primo, 2008). In a study analysing the relative utility of four different types of formative assessment prompts in eliciting students' conceptual understanding, the latter authors found that prompts requiring written responses have the potential to support student understanding of scientific content and processes (Furtak & Ruiz-Primo, 2008).

In summary, various definitions are used in the context of analysing students' development of explanations, partly having similarities to definitions used in the field of argumentation or the construction of models. Consequently, the aim of the analysis also varies with regard to the underlying function of the explanation, i.e. to persuade others or to elaborate on one's understanding. Studies with an explicit focus on explanations often separate content and structure of explanations in the analyses while studies related to argumentation mainly focus on the structural aspects of the explanation. Regarding the format of assessment, students' discourse and written answers to open-ended questions are the dominant data sources in the analyses. With respect to instructional approaches to foster the quality of students' explanations, several studies suggest prompting students to incorporate structural features into their explanations, e.g. by providing groundings or evidence for arguments and claims.

Constructing models

The activity of constructing and using models is addressed in 14 studies in this review. With the exception of one study that simply states that students had time for 'building models' (Wong & Day, 2009, p. 629), all studies provided some further insights into the operationalisation of their understanding of the construct. Looking at these studies in more detail, two types of models – real models and mental models – have to be distinguished. Real models are, e.g. used to support students' learning about complex systems (Hmelo, Holton, & Kolodner, 2000) or to allow students to understand the difference between inference and observation (Akerson & Donnelly, 2010). The majority of publications in this review, however, focus on mental models mostly in the general context of scientific reasoning (e.g. Herrenkohl et al., 1999; White & Frederiksen, 1998). In constructing and using mental models, typical student activities include predicting, controlling, explaining, organising, thinking, reasoning, developing arguments and/or transferring concepts to novel situations. However, the studies in this review vary

considerably with respect to the broadness of their approach to modelling and also in the explicitness of their definitions.

A precise and comprehensive definition is given by Schwarz and White (2005) who define models 'as a set of representations, rules, and reasoning structures that allow one to generate predictions and explanations' (p. 166). The authors understand scientific modelling as a process that involves the construction, evaluation and revision of a model, adding an additional aspect called *meta-modelling knowledge*, i.e. students' knowledge about the nature and purpose of scientific models. A similar approach is employed by White and Frederiksen (1998). In their study, students construct an explicit conceptual model (including scientific laws and representations) with the aim that students should 'understand the form and properties of such scientific laws and models, the inquiry process needed for creating them, and the utility of such models for predicting, controlling, and explaining real-world behaviour' (p. 12). Most studies, however, focus on certain aspects of the above-mentioned definition mostly without explicitly defining their construct. Several studies address the use of models as a tool to support students in developing explanations (Herrenkohl et al., 1999; Sampson et al., 2011), whereas others emphasise the role of models in making predictions (Gobert et al., 2010; Repenning, Ioannidou, Luhn, Daetwyler, & Repenning, 2010) or in understanding the relation between different types of variables (Herrenkohl et al., 2011). The representational aspect is investigated in detail, e.g. by Kaberman and Dori (2009) who explicitly define 'modelling skills as the understanding of correct 3D representation of spatial structures of molecules and the ability to transfer between different molecular representations' (p. 601).

The assessment of modelling competence naturally depends on the focus of the study. Students are, e.g. asked to transfer between representations (Kaberman & Dori, 2009), to explain phenomena (Herrenkohl et al., 1999) or to predict and explain patterns in novel situations (Wilson et al., 2010). A specific paper-and-pencil-based modelling test that includes questions concerning students' meta-modelling knowledge has been developed and used by Schwarz and White (2005). The results obtained in these studies show that emphasising modelling in the learning environment increases not only students' modelling skills (Kaberman & Dori, 2009; Sampson et al., 2011) but also leads to learning gains with respect to inquiry skills and conceptual understanding. In their evaluation of an inquiry-oriented physics curriculum, emphasising the different aspects of modelling, Schwarz and White (2005), e.g. found that the approach facilitated a significant improvement in students' understanding of modelling, and especially meta-modelling, which transferred to inquiry skills and to the learning of science content. Modelling, however, did not only support inquiry learning but also supported students' learning and application of scientific models (White & Frederiksen, 1998; Wilson et al., 2010).

In contrast to most other inquiry activities, almost all reviewed studies with a focus on students' construction of models provide details about the understanding of the construct underlying their investigation. Across the studies, both real models and mental models are addressed, albeit with different aims. While real models are used in the context of complex systems or to emphasise the difference between inference and observation (Akerson & Donnelly, 2010), mental models are used in the context of scientific reasoning (e.g. Herrenkohl et al., 1999; White & Frederiksen, 1998), often with a focus on specific activities, e.g. making predictions or illustrating the relation between different types of variables. Regardless of the type of model, mainly paper and pencil tests with multiple-choice, constructed response or open-ended items are used to assess students' abilities in the activity of constructing models. To support students in constructing models, most studies advocate emphasising modelling in general or different aspects of modelling in the learning environment.

Engaging in argumentation and reasoning

Over the past decade, the study of argumentation has been a prominent feature within research in science education (Osborne, Simon, Christodoulou, Howell-Richardson, & Richardson, 2013). It has been pointed out, however, that researchers often fail to define what exactly they mean by argumentation or argument (Ryu & Sandoval, 2012) and that no consistent usage of the term argumentation has been established – sometimes, it refers to any kind of discussion, sometimes to advancing and evaluating

Table 1. Descriptive characteristics of the publications in the review related to argumentation and reasoning (the numbers give the numbers of publications within the different categories out of a total of $n = 40$ studies focusing on argumentation and providing conceptual details).

Constructs					
Argument(ation) + reasoning 31	Argument(ation) 7	Reasoning 2			
Definition of argument(ation)					
Explicit 27	Implicit 5	Examples 1	Not given 5		
Definition of reasoning					
Explicit 16	Implicit 7	Examples 2	Not given 8		
Context ^a					
Science content 33	Socioscientific/ill-structured 7	Not described 2			
Framework for analysis ^a					
Toulmin ^b 22	Mercer 1	Raven's test ^c 1	Self-developed 13	Content 1	Not described 3

^aStudies can belong to more than one category.^bThe category also includes adapted versions of the original model.^cRaven's test = Raven's progressive tests of non-verbal reasoning.

knowledge claims based on evidence (Shemwell & Furtak, 2010). This inconsistency can also be observed for the publications analysed in this review (for descriptive characteristics of the studies, see Table 1). In total, students' engagement in argumentation and reasoning is addressed in 50 publications – 40 publications provide further details on the operationalisation of this activity. Studies focus on argumentation and/or reasoning, either exclusively or as one important aspect of scientific inquiry; in more than three-quarters of these publications, both constructs are considered together. With regard to the operationalisation, the majority of studies provides details explicitly (or at least implicitly) – nevertheless, details are missing in five publications for argument/argumentation and in eight publications for reasoning, respectively. Among the studies that provide details about the operationalisation of their constructs, differences as well as similarities exist.

The most significant communality can be seen in the importance given to the use of evidence. Almost all studies stress the need to justify different kinds of claims with data or evidence as, e.g. in Clark and Sampson (2007): 'Argumentation includes any dialogue that addresses the coordination of evidence and theory to support or refute an explanatory conclusion, model, or prediction' (p. 255). Nevertheless, a considerable variety among the studies exists with respect to the operationalisation and the analysis of the constructs. Important aspects in this context are, e.g. the differentiation between argument and argumentation, the delimitation to related constructs like explanation and discussion and the frameworks used for defining and analysing argumentation and/or reasoning.

The majority of studies in the review do not systematically differentiate between argument and argumentation – in some cases, both terms even seem to be used more or less synonymously (e.g. in McNeill, 2009). However, there are some studies that explicitly address the differentiation. In these studies, the term argument refers to a product (and the content and structure of this product), whereas the term argumentation refers to a process: 'The former [argument] we see as a referent to the claim, data, warrants, and backings that form the substance or content of an argument. The latter [argumentation], in contrast, we see as a referent to the process of arguing' (Osborne, Erduran, & Simon, 2004, p. 998); a similar differentiation exists in, e.g. Lin and Mintzes (2010) or Ruiz-Primo, Li, Tsai and Schneider (2010). A specific variation of this differentiation can be found in Wilson et al. (2010) as well as in Berland (2011) who define argument as a scientific explanation and argumentation as the process of developing and evaluating such explanations. In accordance with earlier findings (Berland & Reiser, 2009), the studies in

this review show some disagreement whether explanation and argumentation are regarded as complementary (e.g. Berland & Reiser, 2009; McNeill, 2011; Ruiz-Primo, Furtak, Ayala, Yin, & Shavelson, 2010) or as a single practice (e.g. Gotwals & Songer, 2010; Wilson et al., 2010). Berland and Reiser (2009) suggest combining argumentation and explanation into the single practice of *constructing and defending scientific explanations*. This combination in some respect closes the gap to more complementary viewpoints as, e.g. expressed by McNeill (2009): ‘Constructing an explanation does not necessitate using evidence to support a conclusion or trying to convince or persuade another individual that your explanation is correct, yet these are key aspects of scientific argumentation’ (p. 235). The aspect of persuasion is also a key aspect in delimitating argumentation from discussion: ‘The label and term *argumentation* rather than *discussion* was used to emphasise debate and negotiation using specific methods of persuasion’ (Hickey, Taasobshirazi, & Cross, 2012, p. 1255; see also Shemwell & Furtak, 2010).

With respect to argumentation, two major operationalisations can be identified: argumentation as students’ general use of evidence (data and scientific concepts) to construct arguments or explanations about the phenomenon under study (e.g. Erduran, Simon, & Osborne, 2004; McNeill, 2011; Osborne et al., 2004) and argumentation as a social and dialogic interaction in which the participants try to persuade or convince each other of the validity of their claims until one participant (or side) *wins* and the other *loses* (e.g. Berland & Reiser, 2009; Chin & Osborne, 2010). The aspect of persuasion is specifically addressed in approximately one-quarter of the publications in this review. However, several authors stress the point that the social and collaborative component of argumentation is not solely competitive but also a means to collaboratively make sense of the phenomenon under study (Belland et al., 2011) as well as to solve problems and to advance knowledge (Clark & Sampson, 2007). Sampson et al. (2011) even argue that

in science, argumentation is not a heated exchange between rivals that results in winners and losers or an effort to reach a mutually beneficial compromise; rather it is a form of “logical discourse whose goal is to tease out the relationship between ideas and evidence”. (p. 218)

To define, analyse and evaluate argumentation, the majority of studies in the review refers to the model by Toulmin (1958) or adapted versions of his model. The model is used to analyse the structural features and content of arguments produced by single individuals (e.g. Clark & Sampson, 2007; Kelly, Druker, & Chen, 1998; McNeill, 2011) as well as the quality of argumentation in small group discussions. Erduran et al. (2004), e.g. developed a framework where the quality of argumentation is assessed in terms of five levels which illustrate the quality of opposition or rebuttals in the student discussions (Osborne et al., 2004). Despite the prevalent usage of Toulmin’s model in the analysis of argumentation in science classrooms, however, problems can still be observed with respect to the clarification of what counts as claim, data, warrant and backings (Erduran et al., 2004; Shemwell & Furtak, 2010). Some authors thus collapse Toulmin’s data, warrants and backings into a single code called grounds to address the practical difficulty to reliably differentiate among these components (e.g. Clark & Sampson, 2007; Erduran et al., 2004). Shemwell and Furtak (2010) argue that in the field of scientific argumentation, numerous studies can be identified that do not use any normative criteria for what can count as support for arguments. As a consequence, little to no information is provided about the role of students’ subject matter conceptions in their use of evidence or the degree to which students’ arguments reflect scientific criteria for validity.

Among the studies using Toulmin’s model, reasoning is mostly understood as one component of an argument, namely: as the justification that shows why the data count as evidence to support the claim (e.g. McNeill, 2009). Some authors extend this definition by arguing that reasoning should also include the conceptual knowledge that the students apply to a specific situation (e.g. Ruiz-Primo et al., 2010). Other operationalisations define reasoning in a more general way as the process of constructing (White & Frederiksen, 1998) and/or critiquing arguments (Dawson & Venville, 2009; Osborne et al., 2013). However, Toth et al. (2002) focus specifically on one reasoning skill, namely: the evaluation of empirical evidence against multiple hypotheses. In 14 publications, the relationship between argumentation and reasoning remains vague or unclear.

In addition to Toulmin's model, studies in this review use the framework by Mercer (Chin & Teou, 2009); in almost one-third of the publications, self-developed frameworks form the basis of the analysis. These frameworks differ considerably with respect to the comprehensiveness and focus of their operationalisations of argumentation; some studies simply make use of established instruments like, e.g. Raven's matrices for non-verbal reasoning (Osborne et al., 2013) or exclusively focus on whether evidence is provided or not (Gobert et al., 2010). Others, however, use detailed frameworks like the evidence-based reasoning framework (e.g. Brown, Nagashima, Fu, Timms, & Wilson, 2010), frameworks that show similarities to Toulmin without an explicit reference (e.g. Belland et al., 2011; Hickey et al., 2012) and self-developed rubrics, e.g. related to students' ability to defend arguments (Ebenezer et al., 2011). Studies that specifically address the aspect of social interaction within the construct of scientific argumentation use additional coding schemes that help identify the features of the interaction and the nature of the engagement between students (e.g. Clark & Sampson, 2008; Kim & Song, 2006; Osborne et al., 2004; Sampson et al., 2011). Sampson et al. (2011), e.g. coded students' reaction to ideas (accept, discuss, reject and/or ignore) and the overall nature or function of the contributions the students made to the conversation when discussing the merits of an idea (information seeking, expository, oppositional and supportive).

Studies differ not only with regard to the operationalisation of argumentation, but also with respect to the different methods used to assess students' abilities in argumentation. Principally, three formats can be distinguished: transcripts of verbal data of students' discourse (e.g. Osborne et al., 2004); different types of students' written argumentation like, e.g. notebooks (e.g. Ruiz-Primo et al., 2010), evidence maps (Toth et al., 2002) and online discussions (e.g. Clark & Sampson, 2007); and assessment tasks consisting of open (e.g. Lin & Mintzes, 2010; McNeill, 2009), constructed response (e.g. Wilson et al., 2010) or even multiple-choice items (Rivet & Kastens, 2012).

A major difficulty in analysing students' argumentation is the differentiation between the structure and components of an argument and its accuracy. This aspect reflects the above-mentioned question whether scientific argumentation ability includes a component of conceptual quality or not – or, as Kelly et al. (1998) put it, whether an argument is substantive or not: 'An argument is considered substantive when knowledge of the actual content is requisite for understanding' (p. 853; see also Clark & Sampson, 2007). Shemwell and Furtak (2010) explicitly refer to this question by differentiating argumentation and scientific argumentation based on the kinds or levels of support that can warrant knowledge claims. Among the publications in this review, both operationalisations of argumentation in science classrooms exist. Some studies focus solely on the structure and structural components of students' arguments, regardless of the accuracy of the science content (e.g. Cross, Taasobshirazi, Hendricks, & Hickey, 2008; Dawson & Venville, 2009; Erduran et al., 2004; Kelly et al., 1998; McNeill, 2011; Osborne et al., 2004). Others, however, include separate codes to address the aspect of conceptual quality. Clark and Sampson (2008), e.g. coded the conceptual quality of a student comment as either non-normative, transitional, normative or nuanced. Similarly, the accuracy of a claim or a scientific explanation is coded as a separate measure by Ruiz-Primo et al. (2010) and Sampson et al. (2011). Brown et al. (2010) as well as Shemwell and Furtak (2010) eventually included codes for conceptual sophistication, specificity and validity (Brown et al., 2010) and conceptual explicitness, respectively (Shemwell & Furtak, 2010).

The diversity in operationalisations is reflected in the aims of studies investigating argumentation in science education. Generally, four major aims can be identified. The first category is comprised of studies analysing student argumentation in a survey-like manner. Findings consistently show that students struggle with providing high-quality arguments. Arguments are not only found to be largely intuitive and emotive (Dawson & Venville, 2009), but they often include unwarranted claims and miss a rational informal reasoning component (Dawson & Venville, 2009; Kelly et al., 1998). If students provide warranted arguments, however, these show considerable complexity and can be described by three dimensions, the argument strategy, the referent in the warrant and the type of warrant (Kelly et al., 1998).

A second set of studies focuses on effects of teachers' instructional practices. No consistent effects on student reasoning could be found in an intervention analysing the effects of a professional development activity aiming to improve teachers' ability to use instructional practices associated with argumentation

in the teaching of science (Osborne et al., 2013). Teachers' use of a curriculum explicitly designed to support students in the construction of scientific arguments to explain phenomena was investigated by McNeill (2009). She found that some teachers tend to simplify the definition of scientific argumentation, resulting in decreased learning gains in terms of students' ability to write scientific arguments.

The majority of studies in this review address the effects of instructional interventions on student argumentation. Different types of interventions can be distinguished. In general, learning environments specifically designed to foster argumentation result in positive effects on students' argumentation ability (McNeill, 2011; Osborne et al., 2004; Ryu & Sandoval, 2012; Sampson et al., 2011). In their analysis at elementary level, Lin and Mintzes (2010) distinguished between high- and low-achieving students. They concluded that an explicit instruction in argumentation is particularly beneficial for high-achieving students, whereas low achievers lag in their ability to master argumentation skills which was partially attributed to a lack of conceptual knowledge. Gotwals and Songer (2010) eventually found that even when students understand the content, they still have difficulties in creating a complete scientific explanation with a claim, sufficient evidence and reasoning. The relationship between learning gains and engagement in scientific argumentation was analysed by Cross et al. (2008). They concluded that the argumentative structures, the quality of these structures and the identities that students take on during collaborative group work are critical in influencing student learning and achievement in science. Studies investigating the effects of inquiry interventions on students' argumentation lead to inconsistent findings. Whereas some studies report positive effects of inquiry-based instructional activities on reasoning and argumentation (Steinberg et al., 2009; Wilson et al., 2010), others report difficulties in engaging students in high-quality argumentation, especially with respect to the need of supporting claims with data, evidence or reasoning (Ruiz-Primo et al., 2010; Taasooobshirazi & Hickey, 2005). A last type of intervention consists of computer-based scaffolds. Belland et al. (2011), e.g. found that such scaffolds are especially beneficial for low-achieving students in helping them construct more coherent arguments.

The last category of studies consists of evaluations of specific assessment methods designed to assess reasoning and argumentation. Examples are, e.g. the Evidence-Based Reasoning Assessment System (Brown et al., 2010) and the analytical framework for assessing argumentation in online science learning environments developed by Clark and Sampson (2008) that allows for analysing the relationships between levels of opposition, discourse moves, use of grounds and conceptual quality. Erduran et al. (2004) present two methodological approaches that significantly extend and improve the use of Toulmin's model for tracing argumentation discourse in science classrooms. An intelligent argumentation assessment system based on machine learning techniques was eventually developed and evaluated by Huang et al. (2011). The results showed that the system is able to determine the argumentation skill level based on the students' arguments while at the same time promoting students' argumentation levels.

In summary, engaging in argumentation and reasoning is the inquiry activity addressed by most articles in this review. Across these studies, some aspects stand out: most studies put a major emphasis on the use of evidence but do not systematically differentiate between argument and argumentation. In general, there is also an overlap between argumentation and explanation and some authors advocate to combine both into the single practice of *constructing and defending scientific explanations* (Berland & Reiser, 2009).

Two major operationalisations can be identified with regard to argumentation: students' general use of evidence (data and scientific concepts) to construct arguments or explanations (e.g. Erduran et al., 2004; McNeill, 2011; Osborne et al., 2004) and a social and dialogic interaction in which the participants try to persuade or convince each other (e.g. Berland & Reiser, 2009; Chin & Osborne, 2010). To analyse and evaluate argumentation, the majority of studies in this review refers to the model by Toulmin (1958) or adapted versions of his model, but often it is not totally clarified what counts as claim, data, warrant and backings. As a consequence, only few studies provide information about the role of students' subject matter conceptions in their use of evidence (as e.g. in Ruiz-Primo et al., 2010), making it difficult to differentiate between the structure and components of an argument and its accuracy.

Overall, four major aims are identified in the studies investigating argumentation: analysing student argumentation in a survey-like manner, analysing the effects of teachers' instructional practices or instructional interventions on student argumentation and the evaluations of specific assessment methods designed to assess reasoning and argumentation. Regarding assessment formats, the analysis of video- or audiotaped material, of written products (notebooks, online discussions, etc.) and written or computer-based tests (multiple choice or constructed response) is used almost equally frequent in the different studies.

Communicating

Communication is not restricted to a specific stage of the inquiry process but constitutes an overarching ability that serves two major purposes, namely: to better understand scientific concepts and procedures and to participate in a scientific community (Ruiz-Primo, Li, Ayala, & Shavelson, 2004). Among the studies in this review, 23 studies address the aspect of scientific communication. In 17 of these publications, details regarding the operationalisation of this aspect of scientific inquiry are provided. Two broad categories of studies can be distinguished, studies that analyse the structure of the interaction in communication processes and studies that focus on the quality of the interaction.

The first aspect is often analysed in the context of argumentation and explanation in which communication is regarded both as a means to construct and articulate understanding and as a form of social interaction. Berland and Reiser (2009), e.g. investigated how students articulate their understanding as one instructional goal of constructing and defending scientific explanations. The learning environment fostered this goal by highlighting the structural elements necessary in the articulation of an understanding and by explicitly structuring the ways in which students articulated their explanations. In the analysis, two styles of communication could be distinguished, the first weaving together claim, evidence and reasoning components, and the second clearly delineating them. A focus on the relationship between claim and evidence and the justification of the own position can also be found in Kim and Song (2006). In their study, student groups interacted with one another in a peer review process similar to conference presentations by scientists. Results showed that the resulting critical discussions proceeded through the four stages of focusing, exchanging, debating and closing. Based on features of constituent stages, the authors distinguished four types of discussion: exchanging information, consensus, coexistence and extension. A similar approach was followed by Sampson et al. (2011) who analysed the nature of students' reactions to ideas proposed by their peers.

Specific types of conversations emphasising communication as a form of social interaction were investigated by two studies in the context of informal formative assessment. Hickey and Zuiker (2012) analysed conversational turns in student-directed feedback conversations with respect to six mutually exclusive and exhaustive categories of increasingly desirable forms of domain-specific interaction: off task, neutral, procedural, factual, argumentation and argumentation beyond the intervention. The authors found that almost one-third of the conversational turns were coded as off task and only one-quarter as argumentation; no argumentation reaching beyond the intervention occurred.

Ruiz-Primo and Furtak (2007) argue for the importance of social processes, i.e. how knowledge is communicated, represented and argued in the context of assessment conversations:

Social processes refer to the frameworks involved in students' scientific communications needed while engaging in scientific inquiry, and can be oral, written, or pictorial. It involves the syntactic, pragmatic, and semantic structures of scientific knowledge claims, their accurate presentation and representation, and the use of diverse forms of discourse and argumentation. (p. 62)

In their study exploring the relationship between teachers' informal formative assessment practices and students' learning, however, they considered the social processes to be by nature embedded in assessment conversations and thus did not specify teacher interventions in this domain.

As another focus, the nature of student-student but also student-scientist communication was analysed in the context of online learning environments. Kubasko, Jones, Tretter, and Andre (2008) investigated students' synchronous (using live video conferences) and asynchronous (using e-mail)

communication with scientists. Students' questions to the scientists were coded according to five categories: inquiry and interpretation questions, personal questions, technology questions, clarification questions and equipment questions. The authors found that in the synchronous treatment group, most of the questions focused on personal questions about the scientist, whereas in the asynchronous group, they were mostly related to the interpretation of data and use of technology.

The second aspect, namely the quality of students' communication, is mainly addressed with respect to the documentation and communication of inquiry activities carried out by the students. According to Ebenezer et al. (2011), scientific communication involves 'the sharing of ideas with respect to research questions, methods, and claims for peer response and evaluation meeting objectivity from a social perspective' (p. 99). It was operationalised by students' ability to write a clear scientific paper with sufficient details so that another researcher could replicate or enhance the methods and procedures. In comparison with other inquiry abilities, the authors found that students reached comparably lower proficiency values for communication.

In a study evaluating the use of students' science notebooks as assessment tools, Ruiz-Primo et al. (2004) considered each notebook entry as a communication instance. According to the purpose, the entries could belong to different types of communication (e.g. defining, interpreting, concluding but also reporting of an experiment or a procedure) and take different forms, e.g. a table or graph to report data (schematic communication) or a simple description of an observation (verbal communication). The quality of the communication found in students' notebooks was coded based on two criteria: (1) Were students' notebook communications appropriate according to the respective scientific genres (e.g. descriptions or definitions as minor or lab reports as major genres)? (2) Did students' communications indicate conceptual and procedural understanding of the content? Results indicated that students' communication skills and understanding were far from the maximum score and did not improve over the course of instruction (Ruiz-Primo et al., 2004).

Samarapungavan et al. (2008) evaluated the effects of guided inquiry on different measures of student learning in kindergarten. Here, communication was defined as children's ability to discuss, reflect upon and summarise what they had learned. The authors assessed this inquiry activity by analysing students' portfolios with respect to their proficiency in communicating about their investigations verbally or through drawings and pictures. They found that most children were rated proficient or highly proficient. A similar focus on verbal and schematic communication could be found in Gobert et al. (2010). They equated communicating an argument with reviewing, summarising and explaining data and developing and using diagrams; students' abilities in communication were assessed by using open-ended questions.

Two studies eventually addressed communication, specifically in the context of peer and self-assessment. In a methodology paper, Chang et al. (2011) described the development and evaluation of a self-assessment Likert scale for learning science that consists of two subscales, one for inquiry and one for communication. The authors defined communication as a 'meaningful process in which the giver transforms the message into signs (oral, written, or action) and passes it to the receiver' (p. 1219). It consists of four facets: (1) Expressing – use of verbal and written language, mathematical signs, graphs and other representations; (2) Evaluating – analyse or judge the rationality of arguments; (3) Responding – adopt suitable actions based on feedback; and (4) Negotiating – reach an agreement through discussion. White and Frederiksen (1998) included *communicating well* in the reflective assessment part of a physics inquiry curriculum. *Communicating well* was defined as students' ability to 'clearly express their ideas to each other or to an audience through writing, diagrams and speaking. Their communication is clear enough to allow others to understand their work and reproduce their research' (p. 25). Students rated their proficiency on a Likert scale. The authors found that the reflective assessment process appeared to improve social interaction (i.e. teamwork and communication), especially for low-achieving students.

In summary, most studies regard communication as a means to either better understand scientific concepts and procedures or to participate in a scientific community. Unlike the inquiry activities reviewed so far, it is considered an overarching ability that is not restricted to a specific stage of the

inquiry process. Similar to argumentation, communication is mostly analysed with a predominant focus on the structure and interaction in the communication process or with an emphasis on the quality of this interaction. In the former case, the analysis is often conducted in the context of argumentation and explanation, indicating an overlap in both theoretical conceptions and empirical investigations regarding these three inquiry activities.

Few studies try to foster students' communication skills by scaffolding or structuring the ways in which students articulate their understanding (Berland & Reiser, 2009), by computer-based scaffolds (Ebenezer et al., 2011) or by reflective assessment (White & Frederiksen, 1998). However, the baseline of empirical results regarding students' communication skills is that students often reach lower proficiency values for communication in comparison to other inquiry activities (Ebenezer et al., 2011) and these values are often far from the maximum score (Ruiz-Primo et al., 2004). Regarding assessment, mainly video transcripts (e.g. to code conversational turns) and written material (e.g. research papers, constructed response items, notebooks or portfolios) are used.

Summary and discussion of findings

The overarching intention of this review was to contribute to a better understanding of the concept of scientific inquiry. Despite the interest that scientific inquiry has received in science education research in the last decades, there still exists disagreement not only about the efficiency of the approach for student learning, but also about its defining features as an instructional approach.

In order to address this issue, we decided to take a rather atomistic approach by first structuring the overall theoretical construct of scientific inquiry into a list of inquiry activities (cf. Bell et al., 2010; Linn et al., 2004; National Research Council, 1996, 2000, 2012; Pedaste et al., 2015) and then analysing the operationalisations of these single inquiry activities in form of their definitions, their implementation in learning environments and interventions and their assessment (cf. Figure 1). The picture which emerged for each activity on the level of the corresponding empirical studies was illustrated in the previous section. The rationale underlying this approach is that an understanding of the whole first requires a thorough understanding of its constituting parts. However, looking at a complex construct like scientific inquiry in such an atomistic way inevitably leads to the question whether the sum of the parts actually fully represents the whole – a discussion, e.g. known from the field of competence-oriented teaching (e.g. Sadler, 2013). In the following, we will thus try to trace the route from the snapshots of these individual inquiry activities back to the theoretical construct by asking the question what we can learn from this review about the commonalities and discrepancies in the operationalisations of the different inquiry activities, about the characteristics of the collocation of inquiry activities as a whole and about the overarching construct of scientific inquiry as it is reflected in empirical research (cf. Figure 1).

The single bits and pieces

Taking on the atomistic perspective and contrasting the pictures of the single inquiry activities as presented in the previous section, the studies differ in various aspects. These can be allocated to formal and methodological differences like study design and assessment or content-related differences like type and aim of the activity.

Formal and methodological aspects

With regard to the research design, the reviewed studies differ in the instructional setting (ranging, e.g. between project-based teaching and single problem-solving tasks, thus also resulting in huge differences regarding time-on-task), the social setting (e.g. collaborative vs. individualised), the educational setting (e.g. out-of-school labs vs. implemented in regular courses, thus including also low- or high-stakes consequences), the mode of operation (e.g. hands-on vs. computer-based), the student activity (e.g. constructive, receptive, manipulative and/or self-evaluative), the assessment methods (e.g. multiple choice, portfolio and/or written essays) and assessment purpose (formative vs. summative). All

these factors interact with the students' activities and cognitive processes during their inquiry in addition to the type, number and sequencing of different inquiry-oriented activities. However, few studies investigated the effect of manipulating one of these factors and these showed inconsistent findings (e.g. Stecher et al., 2000; Toth et al., 2002). Consequently, the question remains, e.g. to what extent the assessment format or the mode of operation impacts the results of evaluating students' abilities in the different inquiry-oriented activities. Across all activities addressed in this review, these questions are barely researched, let alone answered.

The studies in this review, however, do not only show diversity in the study designs but differ also with respect to the explicitness or vagueness of the descriptions of their theoretical background with several studies giving only implicit and vague details. This vagueness has also been observed and criticised in earlier reviews with respect to the definition of scientific inquiry as a holistic concept (Furtak et al., 2012; Schroeder et al., 2007).

In addition, the summative evaluations of each activity in this review also indicate ambiguities between the theoretical conceptions and their implementation in the individual studies. In the case of argumentation, the model by Toulmin (1958) or an adapted version is used in the majority of studies but often it is not clarified what counts as the fundamental components within this model (i.e. claim, data, warrant and/or backings). Hence, beyond the vagueness in the operationalisation of inquiry-oriented activities themselves as discussed above, the implementation of these conceptions in the different studies also varies considerably. Of course this vagueness may be attributed partly to space limitations in research articles, as this is the sole data source for this review, but it seems doubtful whether different authors (and readers) share the same understanding when using the same terms, on all levels ranging from the general construct of scientific inquiry to the details of implementing specific student activities.

Perspectives and foci on scientific inquiry

In addition to these formal and methodological differences among studies in terms of study design and explicitness, they also differ with respect to the type of activities they focus on. The number of research papers reviewed for each activity differs remarkably, ranging from 11 (searching for information) to 50 publications (engaging in argumentation and reasoning; see Figure 3). Accordingly, specific hot spots of empirical research can be identified focusing mainly on the phases of carrying out experiments as well as explaining and evaluating the results while paying less attention to the preparatory phase, i.e. the identification of research questions, the searching for information and the formulation of hypotheses or predictions.

There seems to be no correlation, however, between the number of publications reviewed for a specific activity and the variance observed in the operationalisation of the respective activities. As in the case of *engaging in argumentation and reasoning* or *constructing and using models*, these activities of scientific inquiry have evolved as their own research fields during the last decades. Here, predominant operationalisations can be identified, e.g. Toulmin's model of argumentation. This conceptual saturation might also indicate a certain degree of elaboration within these research fields which is also supported by existing reviews, e.g. the review of the assessment of modelling competence by Nicolaou and Constantinou (2014). However, the authors also point out that the reviewed studies usually differ vastly according to their operationalisation while also addressing only parts of what can be regarded as modelling competence, indicating both diversity and shortcomings regarding research on these activities. This conclusion can be generalised to all activities reviewed in this paper.

When contrasting the operationalisations of particular inquiry activities, differences between the reviewed studies could sometimes be characterised by a product vs. process dichotomy. For instance, studies incorporating the activity of searching for information had either an information focus or a search focus in their evaluation of student performance. Hence, either the contribution of the collected information to the problem-solving process (e.g. Belland et al., 2011) or students' search behaviour was rated (e.g. Toth et al., 2002). Similarly, engaging in argumentation is sometimes analysed with regard to the content and/or structure of the developed argument (e.g. Berland, 2011) or with regard to the process of argumentation (e.g. Osborne et al., 2004).

A similar classification could be used regarding students' actual performance. Here, studies differ in terms of whether students, e.g. conduct their own investigation (cf. Chen & Klahr, 1999), evaluate a given set-up (e.g. Zion et al., 2005) or self-evaluate their own performance (e.g. Chang et al., 2011). Similarly, students occasionally collected and analysed their own data; sometimes, they were provided with the data. With regard to both the product vs. process dichotomy and the degree of independent agency on the students' side, no studies were identified where the different viewpoints actually made a difference. This aspect stands in close relation to the question of a preferable format for the assessment of all or some of these activities and to which extent the results from different assessment formats and activity operationalisations are actually comparable or rather complementary.

The whole and the sum of the parts

With regard to the rather atomistic approach taken in this review, the question is whether this approach can provide any insights into a more holistic perspective on the construct of scientific inquiry. Regarding the diversity and variability of the operationalisations, instructional settings, social and educational settings, the modes of operation, the assessment methods and assessment purposes implemented in the reviewed studies, it is difficult to find commonalities and common themes. However, when broadening the perspective beyond the individual inquiry activities to their collocation as a whole, three aspects stand out that all pertain to relationships: between the different inquiry activities we *atomised* in this review, between doing inquiry and understanding inquiry and between inquiry and science concepts.

On assembling the different parts

While many studies in this review analysed inquiry-oriented activities as distinct aspects, several contributions tried to find indicators illustrating to which extent students can connect the different activities. Herrenkohl et al. (2011) as well as White and Frederiksen (1998) proposed a coherence score as

a measure of how different parts of the thought experiment are related to one another such as how well the experimental design addresses the hypotheses. In past work the coherence score has been the most sensible score for revealing instructional effects. (Herrenkohl et al., 2011, p. 2)

Regarding the disparity between this emphasis on coherence among the different inquiry-oriented activities and the sole implementation of few more or less distinct activities in numerous research articles included in this review, a desideratum for further research on more comprehensive implementations of inquiry-oriented activities can be put forward. While some studies (e.g. Herrenkohl et al., 2011; White & Frederiksen, 1998) analysed coherence on the students' side, i.e. how different parts of the experiment are aligned to one another as described above, no study was identified that explicitly emphasised the coherence on the conceptual side of implementing scientific inquiry in the classroom, including procedural, epistemic and conceptual features of the distinct scientific activities (as distinguished, e.g. by Furtak et al., 2012; Osborne, 2014). This demand for more research on comprehensive implementations of inquiry-oriented activities may sound trivial but the realisation will prove difficult. Despite the large body of research accumulated in this review, it is challenging to extract a coherent sequence of inquiry-oriented activities (in terms of procedural, epistemic and conceptual features). While it might be easy to agree with the type, number and sequence of activities, the theoretical basis for deciding on the conceptual background for each activity, reflecting the associated epistemic perspective, as well as maintaining coherence across the different activities needs to be extended.

The bird's eye view

The question of coherence is not only a problem for researchers and teachers. Students must also acknowledge and appreciate the function and interplay of the different scientific activities. This epistemology is often referred to as students' understanding of the nature of science (NOS; McComas & Olson, 1998; Osborne, Collins, Ratcliffe, Millar, & Duschl, 2003), the nature of scientific knowledge (Lederman, 2006) or the nature of scientific inquiry (NOSI; Schwartz, Lederman, & Lederman, 2008).¹ Almost all

studies included in this review, however, focus mainly on students' performance when working on inquiry-oriented activities. Hence, a striking finding of this review is how seldom the perspective of students on inquiry in general or on specific activities is explicitly addressed and taken into account in the research literature. Few studies addressed the need for postulating alternative hypotheses, the interpretation of conflicting or contradictory findings, the processing of diffuse data or the discussion of quality standards or good scientific practice (e.g. Schwarz & White, 2005; Vellom & Anderson, 1999; White & Frederiksen, 1998). These studies indicate, however, epistemological constraints in students' perception and interpretation of, e.g. the role of hypotheses in the inquiry process (e.g. Kyza, 2009) which underlines the importance of such activities (in addition to clean and structured examples) to develop students' epistemology which in turn allows students to reflect on the scientific activities they encounter (Pickering, 1992).

Incorporating an epistemological perspective on scientific inquiry might be beneficial not only for students' understanding, but also for the teaching of inquiry. Regarding the analysis of the different inquiry activities in this review, a common instructional approach was to repeatedly expose students to these activities (e.g. identifying questions or formulating hypotheses). However, several studies seemed to make more explicit use of the epistemological structure of the particular inquiry activity. For instance, Toth et al. (2002) proposed the use of evidence maps and reflective assessments that encouraged students to link back their data analysis and interpretation to previous steps in the inquiry process, i.e. to their hypotheses and theories. In the case of constructing models, Schwarz and White (2005) as well as White and Frederiksen (1998) proposed to make use of meta-modelling knowledge, i.e. students' knowledge about the nature and purpose of scientific models. To a certain extent, these approaches of metacognitive elements partly overlap with the epistemological features regarding NOS/NOSI. From this perspective, metacognitive and epistemological aspects might support students in understanding the purpose and goals of the different inquiry activities as well as their interrelations and the utility of the whole process for understanding, explaining, controlling and predicting real-world phenomena. Incorporating these metacognitive or epistemological aspects more explicitly into instructional approaches might result in more efficient teaching strategies than mere repetition of specific activities.

Science and scientific inquiry

With regard to argumentation, explanation and communication, it is evident that these activities are not unique to scientific inquiry but represent major areas within and outside of school. In the pedagogical context, these activities transcend all domain borders, ranging from genres in the language arts to mathematical proofs. With regard to the reviewed empirical studies, this generality is also reflected by the application of domain-general theoretical models in these areas, e.g. Toulmin's model of argument patterns or interaction analysis in communication settings. Consequently, some studies focus solely on structural features, e.g. the structure of an argument, while other studies (also or solely) evaluate the accuracy of the components of an argument, i.e. whether the argument is substantive or not (Kelly et al., 1998). Similarly, the majority of studies focusing on formulating hypotheses often evaluate students' answers solely with regard to whether the proposed hypothesis is testable or not. For these and also some of the other activities, the question arises about the role of science knowledge in these inquiry activities. While a generic perspective on specific inquiry activities is certainly of value for fostering their understanding, e.g. to illustrate characteristics of a hypothesis or what counts as an argument, blending content knowledge and inquiry activities is certainly the more expansive goal of incorporating inquiry in science teaching (cf. Duschl & Grandy, 2011). Authentic science is characterised 'as the integration of the social and material aspects of science' and only the integration of both aspects 'allows students to fully understand how and on the basis of what authority knowledge is formed in the scientific community' (Cavagnetto et al., 2010, p. 429). From this perspective, it is therefore questionable to which extent the process of inquiry is discernible from science content knowledge – or to which extent this disjunction is desirable.

Interestingly, in 46–85% of the studies (depending on the inquiry activity, total mean of 67%), the authors included a science achievement test (partly own developments, partly central standardised

tests, with different foci and length) to relate students' achievement to the results of their analysis of a specific or several inquiry activities. The common goal of this type of analysis was to investigate whether fostering students' ability in certain activities would also increase their science knowledge. The inverse question of how students make use of their science knowledge in acquiring and carrying out a specific activity is addressed only in rare cases. For instance, Kaberman and Dori (2009) differentiated science content, thinking level and chemistry understanding levels with regard to students' ability to formulate hypotheses. Similarly, Samarapungavan, Patrick and Mantzicopoulos (2011) focused on students' ability to use science concepts in the generation of research questions. Across the empirical studies reviewed here, however, the relationship between scientific inquiry and substantive science concepts is almost a blind spot. This is somewhat surprising, especially, for instance, with regard to the NGSS (National Research Council, 2012) and its approach of *three-dimensional learning* along practices, cross-cutting concepts and disciplinary core ideas. It seems that the interplay between these dimensions is illuminated to a lesser extent than commonly assumed.

Guided and self-directed inquiry

In instructional contexts, the complex holistic process of scientific inquiry is often intentionally reduced, especially in the case of guided (in contrast to open) forms of inquiry. Based on earlier work of Schwab (1962), Blanchard et al. (2010) distinguished four levels of inquiry – verification, structured, guided and open, respectively – depending on whether the source of the question, the data collection methods and the interpretation of results are given by the teacher or open to the students. Other models describe inquiry instruction not as discrete levels but as a continuum, ranging from little to more learner self-direction, and more to little direction from the teacher or material (National Research Council, 2000). In this notion, guided scientific inquiry teaching can be regarded as representing the continuum of science instruction between the two extremes, traditional, direct instruction and open-ended scientific inquiry, 'where students are guided, through a process of scientific investigation, to particular answers that are known to the teacher' (Furtak, 2006, p. 454).

It was beyond the means of this review to distinguish the different levels of inquiry in the analysis. However, the feature of guidance or scaffolding was frequently included in studies across inquiry-oriented activities. Regarding the demands which inquiry-based approaches pose to students, it becomes apparent that students with different degrees of ability and experience need specific help and support. Scaffolding and guidance can vary on a continuum, from complete learner self-direction, on the one end, to teacher-led instruction, on the other end. In their meta-analysis, Furtak et al. (2012) concluded that teacher-led inquiry lessons seem to have a larger effect on student learning than those that are student led. However, the mechanism for this differential effect remains unclear. It could be the more direct experience of inquiry on the students' side when the learning conditions are more structured or guided by the teacher (in contrast to student-led conditions; Furtak et al., 2012). Alternatively, it might not be the instructional guidance by the teacher itself but the systematic feedback on students' performance that is more frequent and closer aligned with the different activities of the inquiry process when the learning conditions are more teacher led.

Hence, a further review of inquiry-oriented activities might focus on the use and implementation of feedback and its effect on students' learning. This is especially true for those activities for which repeated practice seems the dominant approach to foster students' abilities, for instance, in case of identifying questions, formulating hypotheses or communication. In addition, metacognitive and epistemological knowledge, as sketched in the previous paragraph, could also be considered the other side of the coin of guidance and scaffolding in inquiry, when identifying metacognition with self-scaffolding (Holton & Clarke, 2006). From this perspective, NOS/NOSI and metacognitive aspects of inquiry might enable students to better develop and monitor their own inquiry activities. If this relation holds true, considering NOS/NOSI and metacognitive elements seems mandatory when teaching inquiry. Otherwise, students' understanding and realisation of scientific inquiry activities might remain bound to scaffolds and other instructional triggers.

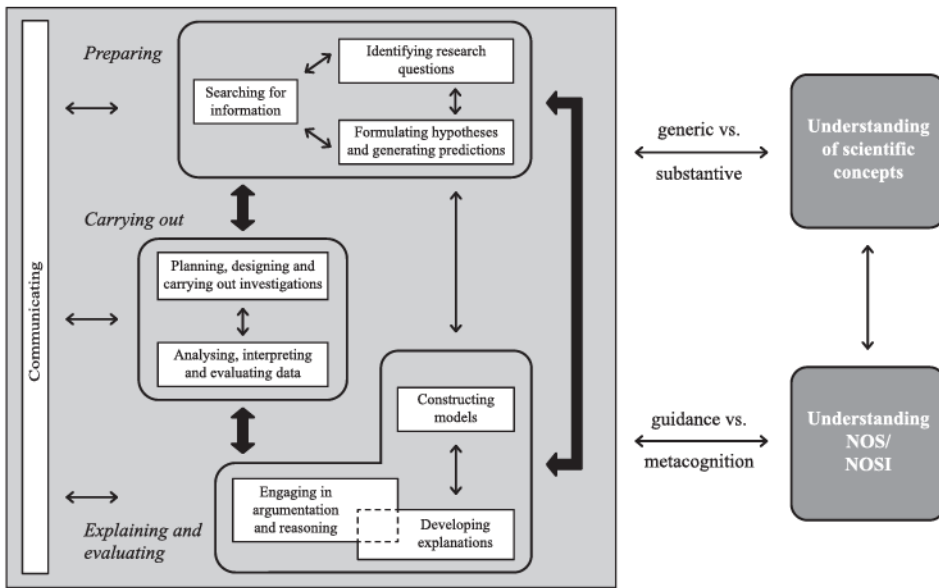


Figure 4. Aggregation of central aspects of the current review. Within scientific inquiry, specific activities (white boxes) are often considered in close correspondence to each other (indicated by black frames) and could be clustered in phases of the inquiry process (preparation, carrying out, explaining and evaluating). These phases are commonly aligned in a circular, interactive sequence (indicated by bolder arrows). Beyond the type, range and sequencing of specific inquiry activities (left side), research about scientific inquiry should be based on a clear theoretical rationale which comprises inquiry, content knowledge and NOS/NOSI.

In total, the detailed analysis of empirical studies both on the level of single inquiry activities as well as on the level of the construct of scientific inquiry has provided complementary insights in this review. When taking a more holistic stance and asking the question how the different aspects could be condensed into a single picture of the construct, it seems obvious that the answer is not just about selecting and sequencing specific inquiry activities. This point is certainly important, but has also been discussed extensively before (cf. Bell et al., 2010; Linn et al., 2004; National Research Council, 1996, 2000, 2012; Pedaste et al., 2015). In summary, specific inquiry activities are often considered closely corresponding to each other (e.g. carrying out an experiment and analysing the obtained data; creating models and developing explanations). These clusters of activities could be subsumed as phases of the inquiry process, including preparation, carrying out, as well as explaining and evaluating (cf. Pedaste et al., 2015). When considering several inquiry activities, these activities (and thereby also the clusters of activities) are commonly aligned in a circular sequence, indicating a reciprocal back and forth between the phases of preparing, carrying out and explaining and evaluating, respectively. Beyond the type, range and sequencing of specific inquiry activities, this review has pointed out that a more holistic picture of scientific inquiry also needs to provide a clear rationale about the relation of scientific inquiry and other fundamental constructs, in particular scientific concepts and knowledge as well as the NOS/NOSI (cf. Figure 4). Regarding the role of scientific concepts, the analytical schemes and rubrics used in the different studies to evaluate students' performance in specific inquiry activities incorporate scientific knowledge sometimes to a greater, sometimes to a lesser extent, i.e. the studies are providing a more generic or more substantive perspective on the inquiry activity. The implications of this shift in the perspective are seldom discussed (Kelly et al., 1998). Most research has focused on using scientific inquiry as a means to foster students' conceptual understanding. Research on how students make use of their conceptual knowledge in inquiry settings, however, seems to be an unexpectedly rare case. There are also few studies that examine the significance of students' understandings of NOS/NOSI in inquiry settings. Here, a more thorough consideration of epistemic aspects in students' inquiry could enable students to better develop and monitor their inquiry activities. So far, this support is mainly

considered in terms of guidance by the teacher or the learning environment, partially also in terms of fostering students' metacognition. However, a more thorough theoretical (cf. Cavagnetto et al., 2010) and empirical consideration of the interrelation between all three constructs could provide an important expansion of the current perspective on research in scientific inquiry, both regarding student performance and the teaching of inquiry.

Limitations

The limitations are generally related to the question of the comprehensiveness of the literature database for the review and thus to the search and selection procedure. A first limitation is given by the selection of keywords for the literature search. Starting from an expansive definition of scientific inquiry, the first step was to generate an initial database that was as comprehensive as possible. By searching in relevant databases, highly visible journals and reference lists of key publications, we sought to find the majority of important contributions to this field of research. Nevertheless, including different or further keywords and considering different or further databases, journals and publications might have led to further relevant publications. A specific aspect of this first limitation is related to the before-mentioned transition from scientific inquiry to scientific practices that started with the publication of the K-12 Framework for Science Education in 2012 (National Research Council, 2012). The scientific practices are closely related to the activities of scientific inquiry that formed the basis of the analyses in this review. Moreover, the time frame of the review ended in 2013. Nevertheless, including the term scientific practices in the keywords might have led to additional entries.

A second limitation is that the sample of reviewed publications was almost exclusively drawn from peer-reviewed, research-oriented journals. This decision was made to ensure a certain level of quality of the reviewed contributions by relying on the journals' policy to ensure a rigorous peer review process. However, focusing on this type of publications may have limited the scope of perspectives on scientific inquiry. Reports, theses or contributions in practice- or teacher-oriented journals may have provided further operationalisations. Regarding the already large number of publications included in this review, however, a review of all inquiry-related publications may not have been achievable. This argument is also related to further decisions made in the literature search, as limiting the sample to contributions published within the last 15 years as well as papers published in English language. In this regard, we tried to be as transparent as possible by explicating the search process.

A third limitation is that this review focuses on students in schools and, thus, takes a certain perspective towards scientific inquiry. As can be seen in Figure 2, numerous studies found by our keyword-based search in databases and journals were excluded because they focused either on students on the tertiary level (19 studies) or on teacher education programmes and teacher professional development (69 studies; cf. Figure 2). Both areas are of course important and relevant but will probably define and operationalise inquiry-oriented activities from a different epistemic and social stance. Further reviews addressing these two areas might provide complementary overviews about research on activities of scientific inquiry with university students and teachers. Contrasting these different perspectives (school vs. university; students vs. teachers) might illustrate interesting changes and transitions in perspectives, but these contrasts were beyond the scope of this review.

Conclusions

This review intended to provide a systematic overview about empirical research on activities that are important constituents of the instructional approach of scientific inquiry. The findings first and foremost illustrate that the variability found in the research literature with respect to the definition and operationalisation of the holistic concept of scientific inquiry is also reflected at the level of single activities of the inquiry process.

Consequently, the research studies accumulated in this review can hardly be condensed to common lines of research, but differ according to numerous factors (e.g. setting, sample or goal). Moreover, the

operationalisations and descriptions of the investigated activities of scientific inquiry as well as the consistency of their implementation differ considerably in depth, comprehensiveness and explicitness among studies. This makes comparisons – and thus the drawing of conclusions regarding the effectiveness of inquiry teaching – difficult, if not impossible (Furtak et al., 2012; Schroeder et al., 2007). Although generally accepted, it seems necessary to repeatedly remind authors of research papers to define and describe the underlying concepts of their studies as comprehensively as possible.

In a similar vein, the interplay between different assessment formats and the obtained results as indicators of students' abilities in the different scientific activities needs to be addressed more specifically. Next to the conceptual discrepancies, the equivalence of results from different data sources is hardly investigated which additionally makes the comparison of findings across studies difficult. The results of this review are thus necessarily more descriptive than explanatory.

Taking the diversity of the compiled findings into account, this review can only be the first step towards a discussion about a more coherent basis of scientific inquiry. The necessity of such a coherent basis is also reflected in the recent shift in terminology from scientific inquiry to scientific practices (National Research Council, 2012). It has been argued that the main problem of teaching science through inquiry has been 'the lack of a commonly accepted understanding of what it means to teach science through inquiry' (National Research Council, 2012, p. 178). The professional practice of teaching science has been undermined by the lack of a clear definition and communication of the activities that students should engage in. Shifting from teaching science as inquiry to teaching science as a practice thus aims to provide a 'greater clarity of goals about what students should experience, what students should learn, and an enhanced professional language for communicating meaning' (National Research Council, 2012, p. 179). However, clarifying the terminology – albeit an important aspect – will not be sufficient to clear up all questions, ambiguities and inconsistencies illustrated in this review.

In closing, the key findings of this literature review can be seen on two levels, the rather atomistic level of the distinct inquiry activities and the more holistic level of scientific inquiry. Regarding the level of the inquiry activities, research on scientific inquiry is a vast field. Conceptual saturation in terms of a predominant model can only be identified for single activities, while research on other activities can mainly be characterised by diversity. Hence, further theoretical work as well as empirical research regarding the interplay of different inquiry activities as well as their individual contribution to the inquiry process as a whole is needed.

From the more holistic perspective, three aspects stand out: first, a case for further research on more comprehensive implementations of inquiry-oriented activities can be made with regard to the sole implementation of few more or less distinct activities in numerous research articles included in this review. Here, both the students' and the researchers' perspectives should be addressed. The impression that conceptualisations of scientific inquiry are an inconsistent accumulation of loosely connected activities might be partly attributed to the rather atomistic approach taken in this review, but it seems doubtful to us whether a more coherent picture would emerge using a more holistic approach.

Second, the interplay between scientific concepts and inquiry activities seems less researched than commonly assumed. While numerous studies included tests on students' conceptual knowledge, the analysis of the interplay is mainly correlational. A more integrated perspective seems necessary to actually understand the role of science in student inquiry.

Third, NOS/NOSI seems to hardly have found its way into research on scientific inquiry, at least when regarding the results of this review that are of course bound to the review's approach (in terms of selected literature databases, searched keywords and analytical methods; cf. Figure 2). Although some studies included measures of students' epistemology or emphasised shortcomings in students' interpretation of the function or purpose of specific inquiry activities, the overall ratio of 5% marginalises the attention which this concept attracts in the reviewed empirical studies. This is to some extent in contrast to the works of, e.g. Sandoval (2003) or Duschl and Grandy (2012) who investigated epistemic aspects in the context of scientific explanation and argumentation. In addition, the aspect of guidance is and has always been a major focus in research (and implementation) of inquiry-oriented activities (cf. Furtak et al., 2012). An important aspect for a future agenda might entail a better understanding of

the interplay and interrelation of different forms of instructional guidance and students' understanding of NOS/NOSI and their *meta-inquiry* knowledge (to revert to the term *meta-modelling knowledge* from Schwarz & White, 2005). In the end, these aspects might be two sides of the same coin: both aim to support students in their inquiry, but the degree of independent agency alternates between instructor and student. Hence, a more systematic consideration of epistemic and metacognitive aspects into the teaching of inquiry might be more beneficial for students' long-term learning and a good indicator to adjust the *fading out* of instructional guidance in the learning process.

Overall and despite its descriptive nature, we believe that the results of this review are valuable to enhance our understanding of scientific inquiry since, unlike previous results, they for the first time allow providing insights into the range of different operationalisations of activities of the inquiry process in empirical research.

Note

1. It is beyond the scope of this review to disentangle an additional unclear concept. The implications, however, to differentiate between characteristics of scientific knowledge and scientific inquiry seem unclear, be it from a theoretical perspective or with regard to their separability in empirical investigations (Neumann, Neumann, & Nehm, 2011). Hence, in the following, the aspects of an 'epistemology of science, science as a way of knowing, or the values and beliefs inherent to scientific knowledge or the development of scientific knowledge' (Lederman, 2006, p. 303) will be subsumed under the acronym NOS/NOSI, treating both as synonymous.

Acknowledgements

The search and procurement of the literature was supported by Hilda Scheuermann and Sabrina Schütz. We thank all partners in the ASSIST-ME project for helpful discussions and comments.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the European Union's Seventh Framework Programme (FP7/2007-2013) [grant number 321428].

Notes on contributors

Silke Rönnebeck (PhD 2001, Kiel University, Germany) has been working as a research scientist at the Leibniz Institute for Science and Mathematics Education (IPN) in Kiel, Germany, since 2004. Her research interests are located at the intersection of assessment and science learning and include, e.g. the assessment of scientific competencies in international large-scale assessments (PISA), formative assessment and inquiry-based learning.

Sascha Bernholt (PhD 2010, University of Oldenburg, Germany) has been working as a research scientist at the Leibniz Institute for Science and Mathematics Education (IPN) in Kiel, Germany, since 2009. His research interests lie on conceptual learning, competency modelling, as well as use of representations in Chemistry Education.

Mathias Ropohl (PhD 2010, University of Duisburg-Essen, Germany) is an assistant professor of Chemistry Education at the Leibniz Institute for Science and Mathematics Education (IPN) located at Kiel University, Germany. In his research, he focuses on the analysis of students' learning in chemistry in the field of scientific inquiry. His aim is to support teachers' use of formative assessment.

References

- Abd El Khalick, F., Boujaoude, S., Duschl, R. A., Lederman, N. G., Mamlok-Naaman, R., Hofstein, A., ... Tuan, H.-L. (2004). Inquiry in science education: International perspectives. *Science Education*, 88, 397–419. doi:10.1002/sce.10118

- Aguilar, O. G., Mortimer, E. F., & Scott, P. (2010). Learning from and responding to students' questions: The authoritative and dialogic tension. *Journal of Research in Science Teaching*, 47, 174–193. doi:10.1002/tea.20315
- Akerson, V., & Donnelly, L. A. (2010). Teaching nature of science to K-2 students: What understandings can they attain? *International Journal of Science Education*, 32, 97–124. doi:10.1080/09500690902717283
- Anderson, R. D. (2002). Reforming science teaching: What research says about inquiry. *Journal of Science Teacher Education*, 13(1), 1–12. doi:10.1023/A:1015171124982
- Bell, T., Urhahne, D., Schanze, S., & Ploetzner, R. (2010). Collaborative inquiry learning: Models, tools, and challenges. *International Journal of Science Education*, 32, 349–377. doi:10.1080/09500690802582241
- Belland, B. R., Glazewski, K. D., & Richardson, J. C. (2011). Problem-based learning and argumentation: Testing a scaffolding framework to support middle school students' creation of evidence-based arguments. *Instructional Science*, 39, 667–694. doi:10.1007/s11251-010-9148-z
- Berland, L. K. (2011). Explaining variation in how classroom communities adapt the practice of scientific argumentation. *Journal of the Learning Sciences*, 20, 625–664. doi:10.1080/10508406.2011.591718
- Berland, L. K., & Reiser, B. J. (2009). Making sense of argumentation and explanation. *Science Education*, 93, 26–55. doi:10.1002/sce.20286
- Blanchard, M. R., Southerland, S. A., Osborne, J. W., Sampson, V. D., Annetta, L. A., & Granger, E. M. (2010). Is inquiry possible in light of accountability?: A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction. *Science Education*, 94, 577–616. doi:10.1002/sce.20390
- Brown, N. J. S., Nagashima, S. O., Fu, A., Timms, M., & Wilson, M. (2010). A framework for analyzing scientific reasoning in assessments. *Educational Assessment*, 15, 142–174. doi:10.1080/10627197.2010.530551
- Butler, K. A., & Lumpe, A. (2008). Student use of scaffolding software: Relationships with motivation and conceptual understanding. *Journal of Science Education and Technology*, 17, 427–436. doi:10.1007/s10956-008-9111-9
- Bybee, R. W. (2000). Teaching science as inquiry. In J. Minstrell & E. H. van Zee (Eds.), *Inquiring into inquiry learning and teaching in science* (pp. 21–46). Washington, DC: American Association for the Advancement of Science.
- Bybee, R. W., Taylor, J. A., Gardner, A., van Scotter, P., Powell, J. C., Westbrook, A., & Landes, N. (2006). *The BSCS 5E instructional model: Origins and effectiveness*. Colorado Springs: BSCS. Retrieved from http://bscs.org/sites/default/files/_media/about/downloads/BSCS_5E_Full_Report.pdf
- Cavagnetto, A. (2010). Argument to foster scientific literacy: A Review of argument interventions in K–12 science contexts. *Review of Educational Research*, 80, 336–371. doi:10.3102/0034654310376953
- Cavagnetto, A., Hand, B. M., & Norton-Meier, L. (2010). The nature of elementary student science discourse in the context of the science writing heuristic approach. *International Journal of Science Education*, 32, 427–449. doi:10.1080/09500690802627277
- Chang, H.-P., Chen, C.-C., Guo, G.-J., Cheng, Y.-J., Lin, C.-Y., & Jen, T.-H. (2011). The development of a competence scale for learning science: Inquiry and communication. *International Journal of Science and Mathematics Education*, 9, 1213–1233. doi:10.1007/s10763-010-9256-x
- 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, C., & Osborne, J. (2010). Students' questions and discursive interaction: Their impact on argumentation during collaborative group discussions in science. *Journal of Research in Science Teaching*, 47, 883–908. doi:10.1002/tea.20385
- Chin, C., & Teou, L.-Y. (2009). Using concept cartoons in formative assessment: Scaffolding students' argumentation. *International Journal of Science Education*, 31, 1307–1332. doi:10.1080/09500690801953179
- Chiou, C.-K., Hwang, G.-J., & Tseng, J. C. R. (2009). An auto-scoring mechanism for evaluating problem-solving ability in a web-based learning environment. *Computers & Education*, 53, 261–272. doi:10.1016/j.compedu.2009.02.006
- Clark, D. B., & Sampson, V. D. (2007). Personally-seeded discussions to scaffold online argumentation. *International Journal of Science Education*, 29, 253–277. doi:10.1080/09500690600560944
- Clark, D. B., & Sampson, V. (2008). Assessing dialogic argumentation in online environments to relate structure, grounds, and conceptual quality. *Journal of Research in Science Teaching*, 45, 293–321. doi:10.1002/tea.20216
- Cross, D., Taasobshirazi, G., Hendricks, S., & Hickey, D. T. (2008). Argumentation: A strategy for improving achievement and revealing scientific identities. *International Journal of Science Education*, 30, 837–861. doi:10.1080/09500690701411567
- Dawson, V., & Venville, G. J. (2009). High-school students' informal reasoning and argumentation about biotechnology: An indicator of scientific literacy? *International Journal of Science Education*, 31, 1421–1445. doi:10.1080/09500690801992870
- Dori, Y. J. (2003). From nationwide standardized testing to school-based alternative embedded assessment in Israel: Students' performance in the matriculation 2000 project. *Journal of Research in Science Teaching*, 40, 34–52. doi:10.1002/tea.10059
- Duschl, R. A., & Grandy, R. E. (2011). Demarcation in science education – Toward an enhanced view of scientific method. In R. S. Taylor & M. Ferrari (Eds.), *Epistemology and Science Education* (pp. 3–19). New York, NY: Routledge.
- Duschl, R. A., & Grandy, R. (2012). Two views about explicitly teaching nature of science. *Science & Education*, 22, 2109–2139.
- Ebenezer, J., Kaya, O. N., & Ebenezer, D. L. (2011). Engaging students in environmental research projects: Perceptions of fluency with innovative technologies and levels of scientific inquiry abilities. *Journal of Research in Science Teaching*, 48, 94–116. doi:10.1002/tea.20387
- Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's Argument Pattern for studying science discourse. *Science Education*, 88, 915–933. doi:10.1002/sce.20012

- Furtak, E. M. (2006). The problem with answers: An exploration of guided scientific inquiry teaching. *Science Education*, 90, 453–467. doi:10.1002/sce.20130
- Furtak, E. M., & Ruiz-Primo, M. A. (2008). Making students' thinking explicit in writing and discussion: An analysis of formative assessment prompts. *Science Education*, 92, 799–824. doi:10.1002/sce.20270
- 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
- Furtak, E. M., Shavelson, R. J., Shemwell, J. T., & Figueroa, M. (2012). To teach or not to teach through inquiry: Is that the question? In S. M. Carver & J. Shrager (Eds.), *The journey from child to scientist. Integrating cognitive development and the education sciences* (1st ed., pp. 227–244). Washington, DC: American Psychological Association. doi:10.1037/13617-000
- 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
- Gobert, J. D., Pallant, A. R., & Daniels, J. T. (2010). Unpacking inquiry skills from content knowledge in geoscience: A research and development study with implications for assessment design. *International Journal of Learning Technology*, 5, 310–334. doi:10.1504/IJLT.2010.037309
- Gotwals, A. W., & Songer, N. B. (2010). Reasoning up and down a food chain: Using an assessment framework to investigate students' middle knowledge. *Science Education*, 94, 259–281. doi:10.1002/sce.20368
- Haury, D. L. (1993). Teaching science through inquiry. *ERIC/CSME Digest*. Retrieved from <http://files.eric.ed.gov/fulltext/ED359048.pdf>
- Heinz, J., Lipowsky, F., Gröschner, A., & Seidel, T. (2012). *Indicators and instruments in the context of inquiry-based science education*. Münster: Waxmann.
- Herrenkohl, L., Palincsar, A., DeWater, L., & Kawasaki, K. (1999). Developing scientific communities in classrooms: A sociocognitive approach. *Journal of the Learning Sciences*, 8, 451–493. doi:10.1080/10508406.1999.9672076
- Herrenkohl, L. R., Tasker, T., & White, B. (2011). Pedagogical practices to support classroom cultures of scientific inquiry. *Cognition and Instruction*, 29(1), 1–44. doi:10.1080/07370008.2011.534309
- Hickey, D. T., Taasobshirazi, G., & Cross, D. (2012). Assessment as learning: Enhancing discourse, understanding, and achievement in innovative science curricula. *Journal of Research in Science Teaching*, 49, 1240–1270. doi:10.1002/tea.21056
- Hickey, D. T., & Zuiker, S. J. (2012). Multilevel assessment for discourse, understanding, and achievement. *Journal of the Learning Sciences*, 21, 522–582. doi:10.1080/10508406.2011.652320
- Hmelo, C. E., Holton, D. L., & Kolodner, J. L. (2000). Designing to learn about complex systems. *Journal of the Learning Sciences*, 9, 247–298. doi:10.1207/S15327809JLS0903_2
- Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42, 99–107. doi:10.1080/00461520701263368
- Hofstein, A., Navon, O., Kipnis, M., & Mamlok-Naaman, R. (2005). Developing students' ability to ask more and better questions resulting from inquiry-type chemistry laboratories. *Journal of Research in Science Teaching*, 42, 791–806. doi:10.1002/tea.20072
- Holton, D., & Clarke, D. (2006). Scaffolding and metacognition. *International Journal of Mathematical Education in Science and Technology*, 37, 127–143.
- Huang, C.-J., Wang, Y.-W., Huang, T.-H., Chen, Y.-C., Chen, H.-M., & Chang, S.-C. (2011). Performance evaluation of an online argumentation learning assistance agent. *Computers & Education*, 57, 1270–1280. doi:10.1016/j.compedu.2011.01.013
- Jiménez-Aleixandre, M. P., & Erduran, S. (2007). Argumentation in science education: An overview. In S. Erduran & M. P. Jiménez-Aleixandre (Eds.), *Argumentation in science education. Perspectives from classroom-based research* (pp. 3–27). New York, NY: Springer.
- Kaberman, Z., & Dori, Y. J. (2009). Question posing, inquiry, and modeling skills of chemistry students in the case-based computerized laboratory environment. *International Journal of Science and Mathematics Education*, 7, 597–625. doi:10.1007/s10763-007-9118-3
- Kelly, G. J., Druker, S., & Chen, C. (1998). Students' reasoning about electricity: Combining performance assessments with argumentation analysis. *International Journal of Science Education*, 20, 849–871. doi:10.1080/0950069980200707
- Ketelhut, D. J., & Nelson, B. C. (2010). Designing for real-world scientific inquiry in virtual environments. *Educational Research*, 52, 151–167. doi:10.1080/00131881.2010.482741
- Kim, H., & Song, J. (2006). The features of peer argumentation in middle school students' scientific inquiry. *Research in Science Education*, 36, 211–233. doi:10.1007/s11165-005-9005-2
- 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., & 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
- Kubasko, D., Jones, M. G., Tretter, T., & Andre, T. (2008). Is it live or is it memorex? Students' synchronous and asynchronous communication with scientists. *International Journal of Science Education*, 30, 495–514. doi:10.1080/09500690701217220
- Kyza, E. A. (2009). Middle-school students' reasoning about alternative hypotheses in a scaffolded, software-based inquiry investigation. *Cognition and Instruction*, 27, 277–311. doi:10.1080/07370000903221718

- Lavoie, D. R. (1999). Effects of emphasizing hypothetico-predictive reasoning within the science learning cycle on high school student's process skills and conceptual understandings in biology. *Journal of Research in Science Teaching*, 36, 1127–1147.
- Lederman, N. G. (2006). Syntax of nature of science within inquiry and science instruction. In L. B. Flick & N. G. Lederman (Eds.), *Scientific inquiry and nature of science: Implications for teaching, learning and teacher education* (pp. 301–317). Dordrecht: Springer.
- Lin, S.-S., & Mintzes, J. J. (2010). Learning argumentation skills through instruction in socioscientific issues: The effect of ability level. *International Journal of Science and Mathematics Education*, 8, 993–1017. doi:10.1007/s10763-010-9215-6
- Linn, M. C., Davis, E. A., & Bell, P. (Eds.). (2004). *Internet environments for science education*. Mahwah, NJ: Lawrence Erlbaum Associates.
- McComas, W. F., & Olson, J. K. (1998). The nature of science in international science education standard documents. In W. F. McComas (Ed.), *The nature of science in science education: Rationales and strategies* (pp. 41–52). Dordrecht: Kluwer.
- McElhaney, K. W., & Linn, M. C. (2011). Investigations of a complex, realistic task: Intentional, unsystematic, and exhaustive experimenters. *Journal of Research in Science Teaching*, 48, 745–770. doi:10.1002/tea.20423
- McNeill, K. L. (2009). Teachers' use of curriculum to support students in writing scientific arguments to explain phenomena. *Science Education*, 93, 233–268. doi:10.1002/sce.20294
- McNeill, K. L. (2011). Elementary students' views of explanation, argumentation, and evidence, and their abilities to construct arguments over the school year. *Journal of Research in Science Teaching*, 48, 793–823. doi:10.1002/tea.20430
- McNeill, K. L., & Krajcik, J. (2008). Scientific explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning. *Journal of Research in Science Teaching*, 45, 53–78. doi:10.1002/(ISSN)1098-2736
- Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction-what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, 47, 474–496. doi:10.1002/tea.20347
- Mistler Jackson, M., & Songer, N. B. (2000). Student motivation and internet technology: Are students empowered to learn science? *Journal of Research in Science Teaching*, 37, 459–479. doi:10.1002/(ISSN)1098-2736
- Mullis, I. V. A., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011 assessment frameworks*. Boston, MA: TIMSS & PIRLS International Study Center.
- National Research Council. (1996). *National science education standards*. Washington, DC: The National Academies Press.
- National Research Council. (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. Washington, DC: The National Academies Press.
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.
- National Research Council. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press.
- Neumann, I., Neumann, K., & Nehm, R. (2011). Evaluating instrument quality in science education: Rasch-based analyses of a nature of science test. *International Journal of Science Education*, 33, 1373–1405. doi:10.1080/09500693.2010.511297
- Nicolaou, C. T., & Constantinou, C. P. (2014). Assessment of the modeling competence: A systematic review and synthesis of empirical research. *Educational Research Review*, 13, 52–73. doi:10.1016/j.edurev.2014.10.001
- Osborne, J. (2014). Teaching scientific practices: Meeting the challenge of change. *Journal of Science Teacher Education*, 25, 177–196. doi:10.1007/s10972-014-9384-1
- Osborne, J., Collins, S., Ratcliffe, M., Millar, R., & Duschl, R. (2003). What "Ideas about Science" should be taught in school science? A Delphi study of the expert community. *Journal of Research in Science Teaching*, 40, 692–720.
- Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, 41, 994–1020. doi:10.1002/tea.20035
- Osborne, J. F., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? *Science Education*, 95, 627–638.
- Osborne, J., Simon, S., Christodoulou, A., Howell-Richardson, C., & Richardson, K. (2013). Learning to argue: A study of four schools and their attempt to develop the use of argumentation as a common instructional practice and its impact on students. *Journal of Research in Science Teaching*, 50, 315–347. doi:10.1002/tea.21073
- Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, 14, 47–61. doi:10.1016/j.edurev.2015.02.003
- Pickering, A. (Ed.). (1992). *Science as practice and culture*. Chicago, IL: University of Chicago Press.
- Repenning, A., Ioannidou, A., Luhn, L., Daetwyler, C., & Repenning, N. (2010). Mr. Vetro: Assessing a collective simulation framework. *Journal of Interactive Learning Research*, 21, 515–537.
- Rivet, A. E., & Kastens, K. A. (2012). Developing a construct-based assessment to examine students' analogical reasoning around physical models in Earth Science. *Journal of Research in Science Teaching*, 49, 713–743. doi:10.1002/tea.21029
- Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching*, 44, 57–84. doi:10.1002/tea.20163
- Ruiz-Primo, M. A., Furtak, E. M., Ayala, C. C., Yin, Y., & Shavelson, R. J. (2010). Formative assessment, motivation, and science learning. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of Formative Assessment* (pp. 139–158). New York, NY: Routledge.

- Ruiz-Primo, M. A., Li, M., Ayala, C., & Shavelson, R. J. (2004). Evaluating students' science notebooks as an assessment tool. *International Journal of Science Education*, 26, 1477–1506. doi:10.1080/0950069042000177299
- Ruiz-Primo, M. A., Li, M., Tsai, S.-P., & Schneider, J. (2010). Testing one premise of scientific inquiry in science classrooms: Examining students' scientific explanations and student learning. *Journal of Research in Science Teaching*, 47, 583–608. doi:10.1002/tea.20356
- Ryu, S., & Sandoval, W. A. (2012). Improvements to elementary children's epistemic understanding from sustained argumentation. *Science Education*, 96, 488–526. doi:10.1002/sce.21006
- Sadler, D. R. (2013). Making competent judgements of competence. In S. Blömeke, O. Zlatkin-Troitschanskaia, C. Kuhn, & J. Fege (Eds.), *Modeling and measuring competencies in higher education* (pp. 13–27). Rotterdam: Sense.
- Samarapungavan, A., Mantzicopoulos, P., & Patrick, H. (2008). Learning science through inquiry in kindergarten. *Science Education*, 92, 868–908. doi:10.1002/sce.20275
- Samarapungavan, A., Patrick, H., & Mantzicopoulos, P. (2011). What kindergarten students learn in inquiry-based science classrooms. *Cognition and Instruction*, 29, 416–470. doi:10.1080/07370008.2011.608027
- Sampson, V., Grooms, J., & Walker, J. P. (2011). Argument-driven inquiry as a way to help students learn how to participate in scientific argumentation and craft written arguments: An exploratory study. *Science Education*, 95, 217–257. doi:10.1002/sce.20421
- Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Sciences*, 12, 5–51.
- Schroeder, C. M., Scott, T. P., Tolson, H., Huang, T.-Y., & Lee, Y.-H. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching*, 44, 1436–1460. doi:10.1002/tea.20212
- Schwab, J. J. (1962). The teaching of science as enquiry. In J. J. Schwab & P. F. Brandwein (Eds.), *The teaching of science* (pp. 3–103). Cambridge: Harvard University Press.
- Schwartz, R., Lederman, N. G., & Lederman, J. S. (2008). *An instrument to assess views of scientific inquiry: The VOSI questionnaire*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Baltimore, MD.
- Schwarz, C. V., & White, B. Y. (2005). Metamodeling knowledge: Developing students' understanding of scientific modeling. *Cognition and Instruction*, 23, 165–205. doi:10.1207/s1532690xci2302_1
- Shemwell, J. T., & Furtak, E. M. (2010). Science classroom discussion as scientific argumentation: A study of conceptually rich (and poor) student talk. *Educational Assessment*, 15, 222–250. doi:10.1080/10627197.2010.530563
- Simons, K. D., & Klein, J. D. (2007). The impact of scaffolding and student achievement levels in a problem-based learning environment. *Instructional Science*, 35, 41–72. doi:10.1007/s11251-006-9002-5
- So, W. W.-M. (2003). Learning science through investigations: An experience with Hong Kong primary school children. *International Journal of Science and Mathematics Education*, 1, 175–200. Retrieved from <http://link.springer.com/article/10.1023%2FB%3A0000016852.19000.af>
- Spires, H. A., Rowe, J. P., Mott, B. W., & Lester, J. C. (2011). Problem solving and game-based learning: Effects of middle grade students' hypothesis testing strategies on learning outcomes. *Journal of Educational Computing Research*, 44, 453–472. doi:10.2190/EC.44.4.e
- Stecher, B. M., Klein, S. P., Solano-Flores, G., McCaffrey, D., Robyn, A., Shavelson, R. J., & Haertel, E. (2000). The effects of content, format, and inquiry level on science performance assessment scores. *Applied Measurement in Education*, 13, 139–160. doi:10.1207/S15324818AME1302_2
- Steinberg, R. N., Cormier, S., & Fernandez, A. (2009). Probing student understanding of scientific thinking in the context of introductory astrophysics. *Physical Review Special Topics – Physics Education Research*, 5, 020104-1–020104-10. doi:10.1103/PhysRevSTPER.5.020104
- Taasoobshirazi, G., & Hickey, D. T. (2005). Promoting argumentative discourse: A design-based implementation and refinement of an astronomy multimedia curriculum, assessment model, and learning environment. *Astronomy Education Review*, 4, 53–70. doi:10.3847/AER2005003
- Taasoobshirazi, G., Zuiker, S. J., Anderson, K. T., & Hickey, D. T. (2006). Enhancing inquiry, understanding, and achievement in an astronomy multimedia learning environment. *Journal of Science Education and Technology*, 15, 383–395. doi:10.1007/s10956-006-9028-0
- Toth, E. E., Suthers, D. D., & Lesgold, A. M. (2002). "Mapping to know": The effects of representational guidance and reflective assessment on scientific inquiry. *Science Education*, 86, 264–286. doi:10.1002/sce.10004
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Tsai, P.-S., Hwang, G.-J., Tsai, C.-C., Hung, C.-M., & Huang, I. (2012). An electronic library-based learning environment for supporting web-based problem-solving activities. *Educational Technology & Society*, 15, 252–264.
- Valanides, N., & Angeli, C. (2008). Distributed cognition in a sixth-grade classroom. *Journal of Research on Technology in Education*, 40, 309–336. doi:10.1080/15391523.2008.10782510
- Vellom, R. P., & Anderson, C. W. (1999). Reasoning about data in middle school science. *Journal of Research in Science Teaching*, 36, 179–199.
- White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, 16, 3–118. doi:10.1207/s1532690xci1601_2

- Wilson, C. D., Taylor, J. A., Kowalski, S. M., & Carlson, J. (2010). The relative effects and equity of inquiry-based and commonplace science teaching on students' knowledge, reasoning, and argumentation. *Journal of Research in Science Teaching*, 47, 276–301. doi:[10.1002/tea.20329](https://doi.org/10.1002/tea.20329)
- Wong, K. K. H., & Day, J. R. (2009). A comparative study of problem-based and lecture-based learning in junior secondary school science. *Research in Science Education*, 39, 625–642. doi:[10.1007/s11165-008-9096-7](https://doi.org/10.1007/s11165-008-9096-7)
- Yoon, C. H. (2009). Self-regulated learning and instructional factors in the scientific inquiry of scientifically gifted Korean middle school students. *Gifted Child Quarterly*, 53, 203–216. doi:[10.1177/0016986209334961](https://doi.org/10.1177/0016986209334961)
- Zion, M., Michalsky, T., & Mevarech, Z. R. (2005). The effects of metacognitive instruction embedded within an asynchronous learning network on scientific inquiry skills. *International Journal of Science Education*, 27, 957–983. doi:[10.1080/09500690500068626](https://doi.org/10.1080/09500690500068626)