Available online at: https://jurnal-fkip.ut.ac.id/index.php/ijrse/issue/archive
International Journal of Research in STEM Education (IJRSE)
ISSN 2721-2904 (online)
Volume 7 Number 1 (2025): 71 - 91

A Systematic Review on Assessment in Inquiry-Based Science Education

[1]Johnson Enero Upahi, [2]Umesh Ramnarain, [3]Ayodele Abosede Ogegbo, & [4]dHsin-Kai Wu
[1]Department of Science Education, Faculty of Education, University of Ilorin, Ilorin, Nigeria
[1,2]Department of Science and Technology Education, Faculty of Education, University of Johannesburg, Auckland Park Kingsway Campus, Johannesburg, South Africa
[4]Graduate Institute of Science Education, National Taiwan Normal University, Taipei, Taiwan

[3]ayo3108@yahoo.com

Abstract

Despite increased advocacy for the use of inquiry-based learning as part of innovative science teaching in various countries over the last decade, research on the assessment of inquiry-based instruction in science education has lagged, with few assessments being implemented and validated. Furthermore, there appears to be a lack of systematic assessment grouping in inquiry-based science education. This systematic review examines 53 empirical studies published between 1996 and 2022, guided by specific assessment design principles and coded using the existing inquiry-based learning framework to identify and categorize key features of inquiry assessment tasks. Results show that most studies adopted the National Research Council's inquiry framework and used constructed-response items as the dominant assessment form. It was also discovered that most studies assessed inquiry tasks at the exploration, interpretation, conclusion, experimentation, questioning, hypothesis generation, and communication sub-phase level. Finally, most of the inquiry assessments were administered via paper-based testing. However, some of the studies reviewed also delivered inquiry assessments using other platforms, such as computer-based, laboratory-based, and mobile device inquiry. Educational implications for future research include using performance-based assessment to comprehensively assess students' inquiry skills

Keywords: *Assessment, inquiry-based instruction, science education, systematic review, performance-based assessment, digital platforms*

INTRODUCTION

Inquiry-based learning, a novel pedagogical strategy that involves students actively participating in the construction of knowledge through exploration and searching for answers to scientific questions based on evidence gathered, has been a key component of science education curriculum reform across the globe (Fang et al., 2016; Keselman, 2003; Pedaste et al., 2012; Shivolo & Mokiwa, 2024). Inquiry in science education is structured as practices students should undertake to learn scientific concepts and understand the Nature of Science. However, there has been a recent shift away from inquiry skills and toward practices to emphasize that conducting science investigation requires skill and knowledge that is unique to each practice. These practices are specific to science and engineering. They include developing and using models, asking questions, planning and conducting investigations, analyzing and interpreting data, constructing explanations, using mathematics and computational thinking, engaging in arguments from evidence, and evaluating and communicating results (Next Generation Science Standards, NGGS Lead States, 2013). To enhance students' capacity to explore scientific phenomena, many science education researchers have proposed and

Corresponding author:
Ayodele Abosede Ogegbo, ayo3108@yahoo.com
DOI: https://doi.org/10.33830/ijrse.v7i1.1744

implemented inquiry-based teaching strategies that emphasize developing students' deep scientific understanding rather than having them memorize science facts (e.g. Lederman et al., 2013; Mupira & Ramnarain, 2018; Pedaste et al., 2015). In light of this, research indicates that teaching science education through inquiry-based instruction enable learners to investigate authentic scientific phenomena and develop abilities like framing an investigable question, observation, carrying out investigation, drawing their conclusions, and sharing their findings with peers (Fang et al., 2016; Shivolo & Mokiwa, 2024; Van Uum et al., 2016). In addition, using inquiry-based science instruction promotes learners' higher-order thinking skills, reduces the achievement gap of minority learners, increases motivation, conceptual understanding, critical thinking, science content understanding, positive attitudes towards science and development of learners' practical skills (Lederman et al., 2019; Fang et al., 2016; Ramnarain, 2014).

The growing emphasis on using innovative approaches to improve students' learning processes has broadened our understanding of strategies that can promote inquiry learning in science classrooms. However, one common issue educators face is determining how to assess inquiry learning in science education. It is believed that students are more likely to demonstrate improved learning on assessments that are tightly related to the curriculum and the learning activities when allowed to interact with inquiry tools and participate in classroom activities that are specifically tied to learning goals in inquiry science education (Liu et al., 2010). According to the National Science Education Standards (NSES), "authentic assessment exercises require students to apply scientific knowledge and reasoning to situations similar to those they will encounter outside the classroom, as well as situations that approximate how scientists do their work" (NRC, 1996, p. 76). Studies have argued that assessment of students' outcomes in science education must involve the generation, interpretation, communication and use of data for a specific purpose that is closely aligned with the curriculum material and science content covered (Harlen, 2013; Liu et al., 2010). This implies that how students approach learning may be influenced by the type of assessment activity they are given. Hence, the main components of assessment in science education should reflect the goals of inquiry-based learning in terms of understanding, skills and competencies (Harlen, 2013). To provide educators and researchers with a more organized and better understanding of inquiry-based science education assessment, we reviewed and synthesized studies on inquiry assessments implemented since 1996 to summarise all inquiry assessment tasks and extract their distinctive features and patterns (Liu et al., 2010).

Research Objectives

In recent decades, the integration of inquiry-based learning into science education has gained significant attention globally, aiming to cultivate scientific thinking, reasoning, and skills among learners. While numerous frameworks and instructional models have been developed to support inquiry-based instruction, relatively fewer studies have systematically explored how such inquiry processes are assessed. The need to align assessment tasks with inquiry frameworks is vital for ensuring that learners are evaluated not only on content mastery but also on their inquiry competencies. In light of this need, this systematic review aims to map and synthesize the characteristics of assessment practices in inquiry-based science education by addressing the following objectives:

1. To identify the inquiry frameworks and conceptual models that underpin the design of assessment tasks in inquiry-based science education.
2. To examine the specific phases and sub-phases of inquiry reflected in empirical assessment tasks implemented in school science contexts.
3. To analyze the types and delivery platforms of inquiry-based assessment tasks utilized in empirical studies, and how these approaches align with inquiry learning goals.

Addressing these questions requires knowledge of relevant literature and frameworks to understand the design of inquiry-based assessment tasks in science education.

THEORETICAL FOUNDATION

The Knowledge Integration Framework for Inquiry Assessment

Though inquiry-based instruction is highly valued in science education, it has not been rigorously assessed. Science assessments in national and international tests have traditionally relied heavily on multiple-choice items that mostly require recall of scientific knowledge. These assessments encourage teachers to prioritize memorization techniques and drills over students' critical thinking (Yeh, 2006). In response to the demand for assessment tasks that promote complex reasoning about relevant science content (Harlen, 2013; Liu et al., 2010), many studies have used the knowledge integration theory as a framework to design, analyze, measure, and describe students' science learning and assessment tasks that promote inquiry-based instruction (Linn et al., 2004; Linn & Hsi, 2000). The knowledge integration theory conceptualizes the addition, differentiation, evaluation, and classification of accounts of scientific phenomena, situations, and abstractions (Linn et al., 2004). A fundamental tenet of knowledge integration is that assessment tasks should pose a problem and necessitate the respondents to develop an argument to assess complex thinking involving the science inquiry process. It focuses on how learners deal with multiple and conflicting ideas about scientific phenomena, create new ideas, and make connections between new and existing ideas to achieve a more normative and coherent understanding (Linn et al., 2004; Linn & Hsi, 2000). The four knowledge integration principles incorporated into the design of science curricula units and assessments within the context of inquiry instructions are as follows: (1) making thinking visible, (2) making science accessible, (3) assisting students in learning from one another, and (4) encouraging lifelong learning (Linn & Hsi, 2000; Liu et al., 2010). According to Linn and Eylon (2006), knowledge integration emphasizes assessing student inquiry by eliciting current ideas about scientific phenomena so they can be reexamined; adding new, more potent ideas to promote links and connections among ideas; developing criteria to distinguish among the full repertoire of ideas; and reflecting on and tracking one's learning to sort out ideas and promote valid coherent views.

The framework highlights students' various and frequently opposing ideas when interpreting scientific phenomena (Linn & Hsi, 2000). Students who engage in knowledge integration distinguish between ideas and can use evidence to support their explanations of scientific phenomena, as opposed to students who add new ideas to their existing repertoire. Students gain a better understanding of inquiry when they apply their science knowledge in the classroom and practical situations. Knowledge integration places a strong emphasis on the idea that for students to develop a comprehensive understanding of scientific concepts, they must recognize the essential elements of scientific phenomena and construct an argument to explain complex or contradictory scientific events. Research provides convincing evidence that using the knowledge integration pattern as the main modality for assessing inquiry encourages active learning, assists students in integrating new ideas with existing knowledge, and helps students develop scientific reasoning (Liu et al.,2010; Stone, 2014). The knowledge integration framework could provide a good starting point for identifying how science assessment activities are designed to measure students' inquiry abilities and reflect the different features of scientific knowledge over tasks that measure conceptual understanding and higher-order level skills in order to provide a clear picture synthesizing inquiry assessment. The knowledge integration theory provided clarity regarding the appropriate conceptual model for the question of this review, and it was also used as the basis of our initial coding scheme and analytical framework. The suitability of the knowledge integration theory in this study is also attributed to the ability to apply its guiding principles in designing or identifying inquiry assessment tasks.

The Evidence-Centred Design (ECD) and Principled Assessment Design for Inquiry (PADI)

In another study, Haertal et al. (2005) conceptualized the Principled Assessment Design for Inquiry (PADI) system based on an Evidence Centred Design developed by (2003). The PADI system was developed to design patterns for assessing inquiry, task design and assessment delivery. According to Mislevy et al. (2003), the evidence-centred design framework consists of three essential components: student models, evidence models, and task models. More specifically, the student model defines the latent traits or constructs that the assessments are designed to measure in terms of knowledge, traits, and abilities within the context of inquiry. The evidence model determines the level of behaviour or performance corresponding to the student achievement level as defined by the student model. Finally, the task model specifies the task features intended to elicit observable behaviours or procedures for developing assessment tasks aligned with the student and evidence models. Hence, the Evidence Centered Design framework provides consistent guidelines for construct development, assessment design, and result interpretation. Since “scientific inquiry requires the use of evidence, logic, and imagination in developing explanations about the natural world" (Newman et al., 2004, p.258), Evidence Centered Design offers an overview of how to identify or develop “assessment tasks that elicit evidence (scores) that bears directly on the claims that one wants to make about” students inquiry practices (Shute et al., 2007, p. 6). As a result, more empirical studies favour the use of Evidence Centered Design to create authentic performance tasks that assess students’ scientific inquiry skills, knowledge and abilities than standardized tests (Baker et al., 2016; Clarke-Midura et al., 2012). Thus, Knowledge integration and evidence-centred design were used as a lens to identify and capture assessment task(s) that present evidence of students' abilities based on the core features of inquiry-based learning processes documented in the literature. While earlier researchers have focused on implementing inquiry-based instructions in science education, there seem to be limited studies on how students are assessed within the context of inquiry-based learning. As a result, researchers have become interested in more practical questions about assessing students' skills from inquiry-based instructions (Harlem, 2013; Liu et al., 2010; Pedaste et al., 2015). Thus, this current review identifies, and analyses published studies that report how inquiry-based tasks can be assessed in science education.

RESEARCH METHOD

Procedure for Literature Search

A systematic review of the literature was conducted to identify the phases/sub-phases of inquiry-based learning measured in inquiry assessment. The EBSCO host online database was used to search for relevant articles on inquiry assessment or testing using the following search terms: inquiry OR enquiry AND assessments OR evaluation OR testing. The EBSCO host Library was selected because the online database has a meta-data of over 80,000 academic journals, 6.5 million books and 450,000 conference proceedings compared to the number of resources in Scopus or ISI Web of Science.

The first step in the procedure for literature search was the selection of online databases referenced in the EBSCO host library. The educational databases include the following: Academic Search Complete, E-journals, ERIC, and Education Source. We searched the online databases for relevant articles ranging from January 1996 to September 2022. The search time was set to begin in 1996 following the publication/ implementation of the United States National Science Education Standards (National Research Council, 1996). In addition, the following search criteria were set in the EBSCO host database: (a) full-text available; (b) references available (c) scholarly (peer-reviewed) journal articles.

Our search generated 919 articles. However, a large number of the retrieved articles did not focus on inquiry-based learning or assessment within K-12 settings. By eliminating the articles that were irrelevant to the study, our search results narrowed to 92 articles. In addition, we applied inclusion and exclusion criteria to identify articles that focused on inquiry assessments. Articles were excluded from the analysis if the focus is on: (a) teacher education and professional development; (b) activities/modules/technology design without evaluation; (c) college students, informal learning, special education and higher education; (d) social sciences, nutrition education, medical sciences, nutrition education and language; (e) secondary analyses of data from PISA and TIMSS, literature review and position papers. The articles that met the inclusion selection criteria were those (a) informed by the practical guidelines on assessment design principles discussed in the introduction, which also featured elements of authentic inquiry-based learning processes, included assessments, (b) focused on science learning of inquiry phases (reasoning, skills, abilities, practices, competencies and thinking) and (c) inquiry-based learning designed for K-12 learners. A snowballing technique was used to screen the reference list of eligible articles that were potentially relevant but were not retrieved initially using the search procedure. This reduced the total number of articles reviewed in this study to 53. A summary of the article selection process for this study is presented in Figure 1.

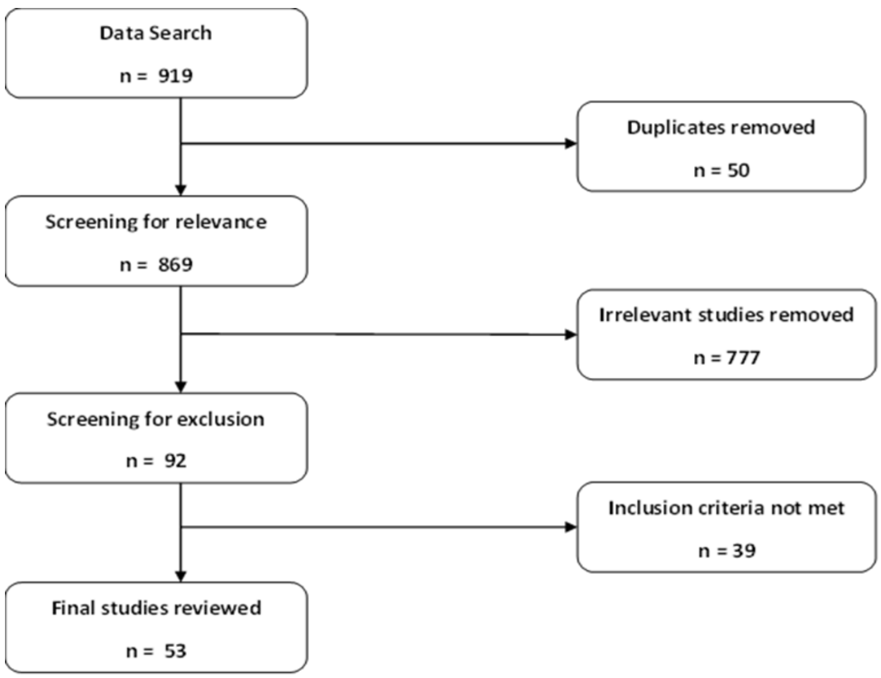

Figure 1: Flow diagram for the article selection process.

Data Extraction and Analysis

Following the selection of articles, all 53 included studies were imported into ATLAS ti, coded, and analyzed for inquiry assessment by categorizing the texts systematically into three stages using content analysis procedures (Fraenkel et al., 2015). ATLAS.ti 8 is a qualitative computer software package that assists researchers in managing textual, graphical, audio, and video data. The steps involved in ATLAS.ti is similar to those used traditionally to analyze text such as notes, documents, or interview transcripts such as preparation, coding, analysis, and reporting. The software offers qualitative researchers two levels of interaction. First, it allows basic "code and retrieval" of data at the textual level. Second, it enables more sophisticated analysis of model-building activities, such as linking codes to form semantic networks and algorithms at the conceptual level.

Using the content analysis procedure, the first author created a coding framework to extract information from each selected article in accordance with the inclusion criteria and the five research questions, as outlined in Table 1. The coding framework emerged from the synthesis of the literature on the core features of inquiry-based learning processes and integrates different observable behaviours as evidence of inquiry learning. Although there are existing inquiry cycles (e.g., de Jong, 2006; Leijen et al., 2012), the high number of different terms used to describe inquiry activities in the literature was informed by Pedaste et al.'s (2015) inquiry-based learning framework which summarises the core phases and processes of inquiry-based learning from learners' viewpoint for teachers and instructional designers. Following that, the second and fourth authors reviewed the coded articles chosen randomly using the initial coding framework. During the review, the authors met with the first author to discuss discrepancies in the codes and improved the initial coding scheme by modifying and clarifying categories, as well as conducting additional reviews of the disputed studies. Finally, after the coding scheme became stable, the authors achieved an inter-rater agreement of 81.9% across all categories, and the first author independently coded the remaining articles. The authors identified overlapping inquiry-learning processes and organized them into a broad framework that sits well with the most common understandings about inquiry-based learning. The result was a comprehensive inquiry cycle with five general phases (Orientation, Conceptualization, Investigation, Conclusion, and Discussion) and nine sub-phases (orientation, questioning, hypothesis generation, exploration, experimentation, data interpretation, conclusion, communication, reflection) of inquiry. A more detailed description of the phases and sub-phases of the inquiry-based learning framework can be found in Pedaste et al. (2015). In addition to the general inquiry phases and sub-phases that served as categories and codes, there were other categories with their corresponding codes, as shown in Table 1.

Table 1. Coding framework for analyzing inquiry assessments

Category	Code
Framework	Inquiry framework/definition
Inquiry/Inquiry Assessment	Reason for assessment development
	Assessment framework
	Assessment Phases/sub-phases
	o Orientation
	o Conceptualization
	▪ Questioning
	▪ Hypothesis generation
	o Investigation
	▪ Exploration, planning
	▪ Experimentation
	▪ Data interpretation
	o Conclusion
	o Discussion
	▪ Communication
	▪ Reflection
	o Future-oriented stage
	o Other
Inquiry/Assessment delivery platform	Paper-based assessment
	Computer-based assessment
	Mobile device
	Others (e.g. equipment and chemicals in a laboratory setting)
Item types	Selected response
	Constructed response
	Performance task
	Other

The first category is the inquiry framework. We analyzed and coded articles for inquiry frameworks that guided researchers in enacting inquiry assessments. Studies that included inquiry definitions/frameworks based on policy documents such as the US NSES, Inquiry and NSES, science practices (NRC, 1996, 2000, 2012) and research-developed frameworks were coded as inquiry frameworks. The second code category concerns the phases/sub-phases of inquiry-based learning/inquiry assessments. This category analyses articles for the five general phases and nine sub-phases of inquiry summarised by Pedaste et al. (2015). Articles were examined for the Orientation phase/sub-phase and coded as such, if the assessment stimulated learners' curiosity about a topic or learning challenge through a problem statement. The Conceptualization phase has two phases: questioning and hypothesis generation. Inquiry assessments that prompted learners to generate scientific question(s) or provided them with investigable questions to guide inquiry tasks were coded into the Questioning sub-phase. Assessments with items requesting learners to formulate testable hypotheses were coded as Hypothesis Generation. The Investigation phase has three sub-phases of exploration, experimentation and data interpretation. Inquiry assessments assigned into the Exploration sub-phase contain items that either provided or required learners to design or develop an action plan in carrying out an experiment/investigation in inquiry tasks. The Experimentation sub-phase focuses on the application of a designed strategic plan for the experiment/inquiry activities. In the Data Interpretation sub-phase, articles are examined if an inquiry task requires learners to make sense of the data collected and/or to synthesize new knowledge from the inquiry process. The Conclusion phase/sub-phase as a code requires students to compare their inference(s) drawn from data with research hypotheses/questions. The Discussion phase has two sub-phases: communication and reflection. In the Communication sub-phase, students present and communicate their findings for comments and feedback from their peers or teachers. An assessment is coded as Reflection if learners are requested to reflect, describe, critique and discuss any of the sub-phases or the whole inquiry cycle or to provide their reasoning with evidence about the phenomenon under investigation.

The third category focused on the item types. This category has three codes: selected response, constructed response and performance task. An inquiry assessment is coded as Selected Response if the assessment requires a forced choice among multiple options, such as in MCQs/tier-response questions. An assessment is coded Constructed Response if learners are requested to provide or generate short or extended answers. Inquiry assessments are coded as Performance Tasks if learners are engaged in hands-on activities to create solutions to questions. The fourth category is the assessment delivery platform. In this category, there are four codes: paper-based testing, web/computer-based programs, mobile devices and laboratory inquiry assessments. If the delivery platform of an inquiry assessment is entirely text-based, it is coded as Paper-based Testing. If an inquiry assessment is delivered through an online learning environment, it is coded as a Web/Computer-based Assessment. Suppose the delivery platform is through an application that can be downloaded from google or iOS play stores and installed on mobile devices or tablets. In that case, the inquiry assessment is coded as a Mobile Device-based Assessment. If an inquiry assessment is based on laboratory activities, it is coded as a Laboratory-based Assessment. For each article, multiple coding was allowed as there could be more than one code in each category. For instance, the inquiry assessment by Fang et al. (2016) included both selected and constructed responses and were coded as such. We coded the articles using the abovementioned coding scheme. After the coding process, we analyzed the data and generated reports of numbers and simple percentages for each category for further analysis. Following this step, the descriptive reports gave us a peek into the pattern of inquiry process(es) reflected in inquiry assessment activities since the publication of the National Science Education Standards (NRC, 1996).

RESULTS

This section presents the findings in the order in which research questions were raised to guide our analysis of inquiry assessment articles. The findings discussed here focused on inquiry definitions/frameworks that guided the inquiry assessments, the process of inquiry within the inquiry assessment tasks, the prevalent assessment types, and the assessment delivery platforms.

What inquiry frameworks are identified from the literature?

In answering the first research question, our analysis identified researchers defined inquiry or the inquiry frameworks they used to assess process(es) of inquiry in their studies. Specifically, we analyzed how researchers operationalized inquiry and the frameworks they adopted or developed to measure students' inquiry abilities. In our analysis, the theoretical/conceptual frameworks used by researchers to foreground their inquiry assessments include, but not limited to the following: Scientific discovery as dual search (SDDS) model proposed by Klahr (2000) (Di Mauro & Furman, 2016; Nehring et al., 2015), science process skills (Prayitno et al., 2017), NGSS standards (Jonathan et al., 2016), inquiry and NSES (McNeill, Pimentel, & Strauss, 2011; Ketlhut et al., 2010; Rhea et al., 2005; Kuo et al., 2015; Wu et al., 2015; Wu, Wu, & Hsu, 2014), abductive inquiry model (Ahmed & Parsons, 2013), inquiry cycle of investigation (Garcia-Milla et al., 2011), claim, evidence and reasoning (CER) model (Bathgate et al., 2015; Wang, 2015), power dynamics + guided inquiry (Donnelly, McGarr, & O'Reilly, 2014), science assessment language demands, SALD (Bunch et al., 2010), Toulmin's model (Huang et al., 2011) inquiry island (Eslinger et al., 2008), argument-driven inquiry (ADI) instructional model (Demircioglu, 2015; Sampson et al., 2013), Toulmin's argument pattern, TAP (Hsu, Chiu, Lin, Wang, 2015), 5E + I/A inquiry model (Zuiker & Whitaker, 2014), cognitive problem-solving process inquiry (Kremer et al., 2014). In the articles analyzed, majority of the research on inquiry assessment defined inquiry or used the different modified inquiry frameworks in policy documents developed by the National Research Council (NRC, 1996, 2000). The inquiry frameworks that guided researchers in enacting inquiry assessment are included in the list of articles selected for this review (see Appendix 1).

What are the phases/sub-phases of inquiry within the assessment tasks?

In answering research question 1, the inquiry phases/sub-phases measured in the reviewed articles were categorized according to the new inquiry-based learning framework proposed by Pedaste et al. (2015). The frequency of the sub-phases of inquiry in the reviewed articles differs significantly. Table 2 presents the sub-phases of exploration and data interpretation (43, 14.7%) as the most frequent processes of inquiry measured in the inquiry assessments, followed by experimentation and conclusion (39, 13.4%). Next were the sub-phases of questioning (34, 11.6%), hypothesis generation (31, 10.6%) and communication (30, 10.3%). The frequency of process(es) involved in getting learners to start an investigation, coded as orientation (14, 4.8%), and the sub-phases of reflection (16, 5.5%) and future-oriented stage (3, 1.0%) were the least measured. The spread of the sub-phases of inquiry process in the articles can be better appreciated in Figure 2. Interpreting the result of our analysis in terms of the general phases of inquiry, the majority of articles focused on investigation and conceptualization phases. With multiple coding, the summation of the number of sub-phases of inquiry assessments does not total the number of inquiry articles.

Table 2: Frequency and percentage distribution of the sub-phases of inquiry assessment

General Phases of Inquiry	Sub-phases of Inquiry Assessment	Frequency	Percentage
Orientation	Orientation	14	4.8
Conceptualization	Questioning	34	11.6

General Phases of Inquiry	Sub-phases of Inquiry Assessment	Frequency	Percentage
	Hypothesis Generation	31	10.6
	Exploration	43	14.7
Investigation	Experimentation	39	13.4
	Interpretation	43	14.7
Conclusion	Conclusion	39	13.4
Discussion	Communication	30	10.3
	Reflection	16	5.5
	Future-oriented Stage	3	1.0
	Total	**292**	**100.00**

Note. *The frequency does not total the number of inquiry articles coded but the occurrence of the sub-phases of inquiry in the assessment tasks in each article.

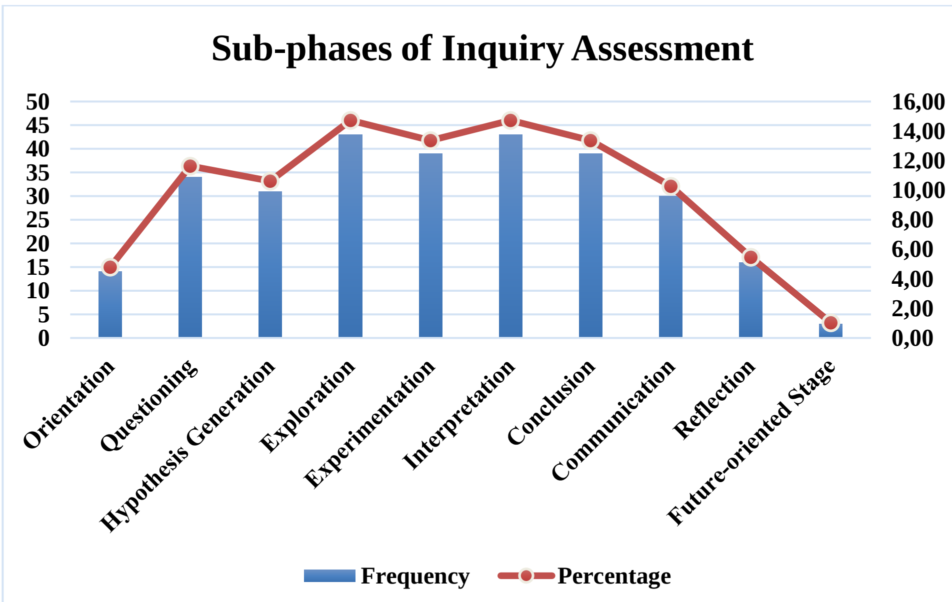

Figure 2. The sub-phases of inquiry assessment identified in the empirical studies examined

One of the requirements for a task or an activity to be considered as an inquiry is the evidence of an investigable question to be explored. In this review, researchers initiated inquiry assessments by providing opportunities for learners to engage in scientifically oriented questions. Learners were prompted to 'pose questions to explain observations' (Day & Matthews, 2008), 'ask relevant questions regarding the phenomena they have observed…' (Hofstein et al., 2004; Rhea et al., 2005), provide research questions (Schwarz & White, 2005), and formulate questions (Kremer et al., 2014). Although the term 'orientation' as a sub-phase of inquiry was not particularly used in any of the articles reviewed, intermediate processes or activities described by (Pedaste et al., 2015), such as 'anchor', 'simple observation', 'introducing a topic' or 'theory' that convey similar meanings were coded as orientation. For instance, in an inquiry task, learners were expected to 'build a connection between their prior knowledge and the current research subject' (Demircioglu & Ucar, 2015) and 'seek information to refine the research questions' (Day & Matthews, 2008). These were to serve as 'anchors', 'exploration' and a way of 'introducing the topic/phenomenon' to be investigated to the learners.

The phases of investigation and conclusion record high frequency in the inquiry assessments. The investigation has two sub-phases of experimentation and data interpretation that involve a simple observation process with respect to the research question and collection of evidence concerning a hypothesis. In this review, we found examples of inquiry tasks that addressed these sub-phases. In developing enquiry skills using position-link datalogging, Davis et al. (2012) engage learners children in a paper-based task to extract information from a bar chart and interpret a line graph of data obtained from Science Scope's Datadisk software. This task draws on a wide range set of interpretation skills. In a more authentic inquiry-based assessment, learners were required to perform tasks that involved planning and carrying out experiments to compare different brands of toothpaste, use chicken eggshell to simulate human teeth since the major component of both an eggshell and human teeth is calcium carbonate, with a goal of investigating the effects of toothpaste on the rate of tooth decay (Cheung, 2005).

What types of assessment are prevalent in the assessment tasks?

In all the 53 identified inquiry research articles, the most common type of assessment task was constructed response (CR) items (46.6%), followed by selected response (SR) items (28.8%) and then performance tasks (PT) items (24.7%). The different types of assessment used in the examined studies are presented in Figure 3. A quick look at Figure 3 indicates researchers' preference for inquiry assessments primarily focused on CR items. While majority of the inquiry assessments required CR from learners, items on SR and PT were relatively small, with a marginal difference between SR and PT items. Inquiry assessments that focused on objective testing (e.g., multiple-choice items), as observed in some of the SR items, may only require learners to recall or apply a few pieces of knowledge to solve inquiry tasks. However, an interesting finding is that some of the SR items were in tier-response formats that require learners to explain or justify their choices. In some studies, the SR items were used with the CR items to strengthen and guarantee the construct validity of inquiry abilities measured in the assessments. The CR and PT reflect item types that engage learners in hands-on activities during inquiry and open-ended questions.

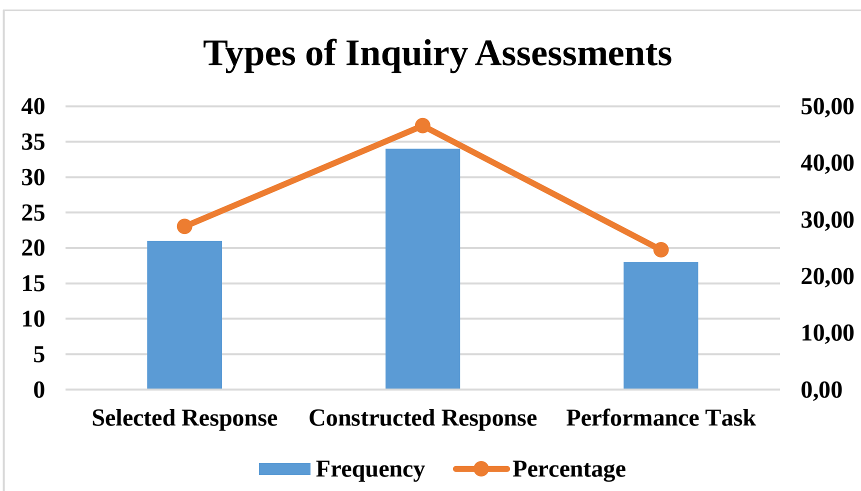

Figure 3. The different types of inquiry assessments used in the examined studies

Surprisingly, our analysis revealed SR items in inquiry assessments, a closer examination of the quotes on SR items revealed that some studies used SR and CR items together to measure inquiry abilities. Although, we identified literature in which researchers administered only multiple-choice questions (MCQs) to assess inquiry abilities with exceptions (Blanchard et al., 2010; Clark & Linn, 2003; Gobert et al., 2011; Nehring et al., 2015; Silk et al., 2009). Although Blanchard et al. (2010) used multiple-choice items that were carefully crafted from VNOS (for NOS questions), the nature of scientific knowledge and science practices, and participants responded in writing to a few open-ended questions about the NOS, followed by in-depth interviews. Rather than solely using force-choice MCQs, participants in the study had the opportunity to reflect on their answers with additional clarifications.

In other reviewed articles, researchers who recognized the limitations of objective testing in revealing the complex thinking students engaged in during inquiry opted for either CR items or SR + CR items in inquiry assessments (Ahmed & Parsons, 2013; Bathgate et al., 2015; Demircioglu & Ucar, 2015; Fang et al., 2016; Kuo et al., 2015; McNeill et al., 2013; Songer et al., 2002; Turkan & Liu, 2012; Vitale et al., 2016). For instance, Demircioglu and Ucar (2015) measured students' inquiry abilities in a quasi-experimental research that is an argument-driven inquiry and based on laboratory instruction. The inquiry assessments administered were SR/MCQs and CR. The written reports were in line with the data collected and the results obtained from experimenting with the phenomenon of geometrical optics. While students' written reports in the experimental group were expected to answer questions on "what they were trying to do and why?", "what did they do and why?" and "what is the argument?". The MCQs on optical achievement and science process skills tests measured content knowledge and inquiry abilities respectively.

What assessment delivery platforms was used for inquiry assessments?

In the articles reviewed, the most common platform used to deliver inquiry assessments was paper-based testing (PBT) followed by computer-based assessment (CBA), laboratory-based assessment (LBA) and mobile device inquiry assessment (MD) in order (PBA = 48.4%, CBA = 34.4%, LBA = 14.1%, MD = 3.1%). Figure 4 presents a graphical view of the assessment delivery platforms for inquiry assessments. While PBT appeared to be the predominant delivery platform used to assess inquiry process, a significant finding is the proportion of web/computer-based assessment (W/CBA) systems that allow learners to engage in an immersive, three-dimensional environment navigating virtual worlds to make observations and gather data to solve scientific problems. In some other articles, the delivery platform were typical laboratory environments, where learners manipulate apparatus and make observations of phenomena under investigation. Although the frequency of LBA was low, inquiry assessments on mobile devices were the least.

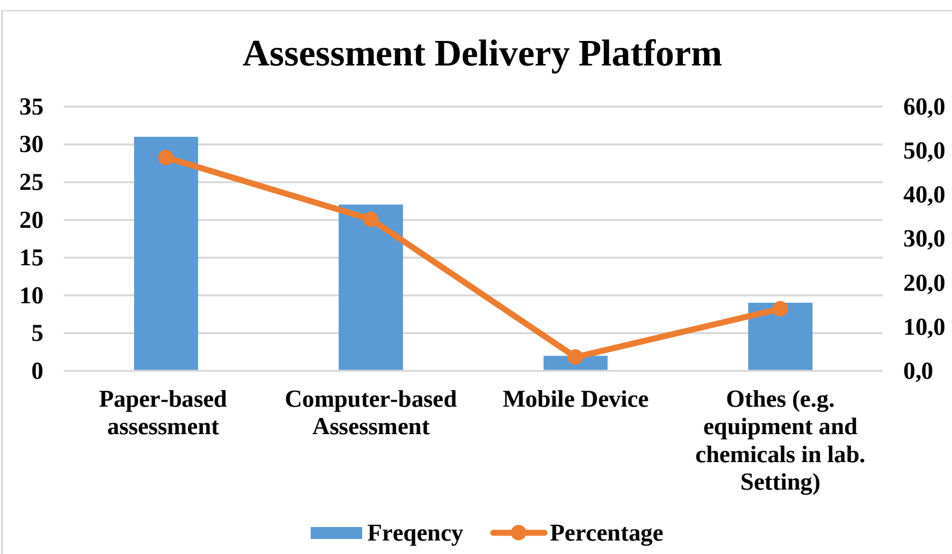

Figure 4: Assessment delivery platform used in the empirical studies selected.

DISCUSSION

Despite the effectiveness and popularity of inquiry-based learning in science education, one of the factors impeding its development in schools is assessment. Assessment is an essential component of teaching and learning because it is used to evaluate the needs of students, instructors, schools, curriculum, and instructional programs and provide the necessary feedback for improvement. For inquiry learning in science to be effectively reinforced in K-12 classrooms, science assessments must reflect an understanding of the scientific concepts taught and align with the goal of inquiry-based learning and the development of scientific practices. Though inquiry in science education places many demands on assessment processes, there are limited resources to meet these demands and numerous challenges to developing satisfactory systems for assessing inquiry science, especially when modifying existing practices (Hein & Lee, 2000). This study systematically reviews and categorizes existing studies on inquiry assessments published between 1996 and 2020. It focuses mainly on Pedaste et al.'s (2015) inquiry-learning framework but also draws on the principles for the inquiry assessment design, which is explained in knowledge integration theory (Linn et al., 2004).

According to findings, most of the studies reviewed mapped out their framework for assessing inquiry learning based on the definition of inquiry in standards documents such as the National Science Education Standards (NRC, 1996; 2000). The National Science Education Standards (NSES) described scientific inquiry as a complex set of activities that involve the use of skills like "observations, posing questions, examining books and other sources of information to see what is already known, planning investigations, reviewing what is already known in light of experimental evidence, using tools to gather, analyze, and interpret data, proposing answers, explanations and predictions, and communicating the results. Inquiry requires identifying assumptions, using critical and logical thinking, and considering alternative explanations." (p 23). The NSES provides a set of outcomes that promote science literacy and defines what students should know, understand, and be able to do in science. As a result, it serves as a model for many other nations to develop their academic standards. However, due to developmental differences in learners' fundamental abilities and understandings of inquiry across ages, several frameworks and models for assessing inquiry learning were conceptualized, all of which are based on the standards (Abdusselam et al., 2018; Ahmed & Parsons, 2012; NGGS Lead State, 2013; Pedaste et al., 2015). Nevertheless, the various framework consistently stresses the importance of asking questions or posing problems, planning and carrying out investigation, mathematical thinking, making inferences, engaging in argument from evidence, analyzing and interpreting data, and effective use of communication skills in explaining facts with evidence when sharing and discussing findings or conclusions as practices needed to engage and assess inquiry (NGSS Lead States, 2013; NRC, 2012; Pedaste et al., 2015). In order to understand how inquiry skills are assessed in science classrooms, findings from the reviewed studies revealed that assessment tasks were administered to students based on different phases/sub-phases of the inquiry-learning framework. However, most of the assessment tasks identified in the studies appeared to focus on the hands-on component of inquiry and ignore the minds-on component. It was also discovered that 14.7% of the studies assessed students' inquiry skills during the exploration (simple observation) and data interpretation phase. In comparison, 13.4% assessed students' inquiry skills during the experimentation and conclusion phases. This implies that most of the sampled studies emphasize the importance of developing students' abilities to explore and interpret data based on information gathered and use such data to make informed decisions. Nevertheless, students' inquiry skills were also assessed in the questioning phase (11.6%), hypothesis generation phase (10.6%), communication phase (10.3%), orientation phase (4.8%), reflection phase (5.5%), and future-oriented stage (1.0%). The low percentage of assessment tasks reviewed during the hypothesis generation phase could be due to the fact that hypothesis generation requires consistent thinking, reasoning, calculation, and planning creativity (Osborne, 2015), making it one of the most challenging aspects of inquiry for students (Kuang et al., 2020). According to Davis and Bellocchi (2018), developing students' capacity to design and conceptualize scientific investigation (orientation) is necessary for facilitating students' inquiry. However, Herranen and Aksela (2019) argue that designing experiments that allow students to create questions is essential in developing students' capacity to do inquiry.

Another main research question that this study answered was to investigate the types of assessment prevalent in the assessment tasks. Emphasis on these types of assessments was necessary for educators to understand the kind of tasks or activities they can use to assess students' inquiry learning in science. In this review, the most common type of assessment task was constructed response (CR) items (46.6%), followed by selected response (SR) items (28.8%) and then performance tasks (PT) items (24.7%). Although this study identified literature in which researchers administered only multiple-choice questions (MCQs) to assess inquiry abilities with exceptions, research has indicated that multiple-choice tests are not a good measure of higher-order and complex skills (Clarke-Midura et al., 2012). The development and demonstration of higher-order cognitive skills involved in scientific inquiry are difficult to measure with open-response and multiple-choice tests (NRC, 2006). Moreover, Ketelhut and Dede (2006) argue that the complex nature of scientific inquiry is better captured using an alternative assessment method in addition to a more traditional multiple-choice test. As a result, Clarke-Midura et al. (2015) claim that using performance assessment tasks in assessing students' inquiry is valuable for aiding learning and providing formative feedback to teachers about ongoing student attainment. Performance-based assessment tasks support students' reasoning within complex problem-solving situations and provide better measures of science inquiry (Darling-Hammond & Adamson, 2010). Research supports that performance tasks are valuable both for aiding learning and for providing formative feedback to teachers about ongoing student attainment. With regards to the assessment delivery platforms used for inquiry assessments, 48% of the reviewed studies used paper-based testing (PBT), 34.4% used computer-based assessment (CBA), 14.1% used laboratory-based assessment (LBA), and 3.1% used mobile device inquiry assessment (MD). Paper-based testing focuses on individual paper-pencil test items and relies on student affirmation as a response that indicates knowledge (Clarke-Midura et al., 2012). Research indicates that the assessment of inquiry-related questions using paper-based testing exhibits a threat to validity and does not provide the authentic context for inquiry as a complex and multifaceted activity involving both cognitive and physical activity (Ramnarain, 2014). This type of assessment has relatively few opportunities to measure the complex scientific knowledge and skills that inquiry instruction intended to target (Quellmalz & Pelligrino, 2009). Research indicates that inquiry skills should be assessed within the scientific context in which they are developed (Mislevy et al., 2003). As a result, it is believed that one of the most effective ways to learn inquiry in science is through laboratory activities (Minner et al., 2020). This could explain the use of laboratory (hands-on or computer-based or mobile device inquiry) assessment in the studies examined. However, the low use of hands-on laboratory environments as platforms for assessing inquiry in the reviewed studies could be attributed to difficulties associated with the physical laboratory environment. These difficulties include the "potential hazards of chemicals, the high cost of laboratory equipment and materials, liabilities of using tools or other laboratory materials, and the use of classroom hours to (repeatedly) set up traditional experiments" (Scalise et al., 2011, p. 1051). In order to address issues with the use of hands-on laboratory environments for assessing inquiry, research recommends the use of virtual laboratory environments made possible by computer technology and mobile devices (de Jong et al., 2013).

A notable finding in this review is the gradual and increasing use of CBA to deliver inquiry assessments. The significance of this is the opportunity immersive technologies afford learners to navigate avatar virtual worlds (virtual and augmented reality) to make observations, collect and analyze data and report their findings in authentic scientific inquiry (e.g., Hsu et al., 205; Kuo et al., 2015; Nelson et al., 2013). While no study used immersive technologies to deliver all the science practices, our review showed inquiry assessments in the disaggregated components of scientific inquiry (e.g., Fang et al., 2016; Nelson et al., 2013; Scalise & Clarke-Midura, 2018). Furthermore, international high-stakes testing programs like Trends in International Mathematics and Science Survey (TIMSS) and Programme for International Student Assessment (PISA) are beginning to deliver assessments through technology (e.g., Mullis et al., 2020; Quellmalz & Pellegrino, 2009), indicating that they all recognize the potential of using technology to assess scientific inquiry. However, using technology to deliver digitized versions of paper-based tests does not realize the full power of technology-enabled (computer-based or mobile device) assessments to create new measurement experiences that provide richer observations of student learning (Clarke-Midura et al., 2012). In addition, Shavelson et al. (1991) indicate that laboratory assessment (hands-on) and computer-based assessment (virtual performance) do not tap the same knowledge. Though the two common delivery platforms used in the reviewed studies are PBT and CBA, it appears that CBA is more effective than PBT in promoting and assessing science content knowledge and inquiry skills.

Limitations

It is acknowledged that systematic review has some limitations. A limitation of this study, as with other review studies, is that while we attempted to cover a broad range of studies on this topic, it is unlikely to include every relevant study. Even though various search phrases were tried to see if any additional relevant studies could be found, the inclusion criteria used in the study may not have captured all the pertinent literature because they were too restrictive. Although inquiry learning is an inclusive instructional practice, it was decided to focus only on specific studies where the terms inquiry, inquiry, assessments, evaluation, and testing in school science were evident or embedded in the research. Therefore, some studies that set out to capture the implementation of inquiry but were not focused on assessment may have been excluded.

CONCLUSIONS

This systematic review analyses assessments in inquiry-based science education and identifies current gaps and future directions for assessing inquiry abilities. According to the findings of the reviewed studies, inquiry assessments were developed and used more frequently in middle and upper grades than in elementary schools. This review paper also identified various inquiry frameworks and models that differ in the elements and indicators that guide the design of inquiry assessment. However, the elements and indicators of inquiry process identified in the various frameworks were based on the different dimensions of scientific inquiry established by the Framework for K-12 Science Education (NRC, 1996; 2012). Moreover, the study findings seem to suggest that inquiry-related tasks assigned to students can be assessed when students: conceptualize an investigation, generate a hypothesis, make predictions or ask questions, make simple observations or collect evidence concerning a hypothesis, interpret data during an investigation, draw conclusions/inferences about their findings, and discuss/communicate their findings. Three types of assessment tasks emerged from the literature: constructed response (CR) items, selected-response (SR) items and performance tasks (PT) items. Though several studies used both CR and SR items to reinforce and ensure the construct validity of the inquiry skills and abilities tested in the assessment tasks, most studies used constructed response (CR) items for subjective assessment. Although most of the studies assessed inquiry tasks using traditional paper and pencil assessment, this method does not provide an adequate and authentic measure of students' understanding of inquiry skills and ability (Quellmalz & Pellegrino, 2009; Ramnarain, 2014). Only a few studies reported using performance-based practical laboratory assessments to evaluate students' inquiry skills. However, there is evidence from the findings that educational researchers and teachers are now using new formats within computer-based performance assessment to encourage students' active participation in inquiry learning and evaluate their inquiry abilities. This shows the importance of using contemporary technologies in science education. The rise of digital networks has increased the use of technology in creating inquiry learning environments, as well as in the improvement of scientific investigations and communications. In order to support the understanding and implementation of inquiry-based science education, more emphasis needs to be placed on using performance-based assessments to evaluate students' inquiry skills and knowledge.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported that there is no funding associated with the work featured in this article

REFERENCES

Abdusselam, M. S., Kilis, S., Şahin Çakır, Ç., & Abdusselam, Z. (2018). Examining microscopic organisms under augmented reality microscope: A 5E learning model lesson. Science Activities, 55(1-2), 68-74. https://doi.org/10.1080/00368121.2018.1517717

Ahmed, S., & Parsons, D. (2013). Abductive science inquiry using mobile devices in the classroom. Computers & Education, 63, 62-72. https://doi.org/10.1016/j.compedu.2012.11.017

Baker, R. S., Clarke-Midura, J., & Ocumpaugh, J. (2016). Towards general models of effective science inquiry in virtual performance assessments. Journal of Computer Assisted Learning, 32(3), 267-280. https://doi.org/10.1111/jcal.12128

Bathgate, M., Crowell, A., Schunn, C., Cannady, M., & Dorph, R. (2015). The learning benefits of being willing and able to engage in scientific argumentation. International Journal of Science Education, 37(10), 1590-1612. https://doi.org/10.1080/09500693.2015.1045958

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(4), 577-616. https://doi.org/10.1002/sce.20390

Bunch, G. C., Shaw, J. M., & Geaney, E. R. (2010). Documenting the language demands of mainstream content-area assessment for English learners: Participant structures, communicative modes and genre in science performance assessments. Language and Education, 24(3), 185-214. https://doi.org/10.1080/09500780903518986

Cheung, D. (2005). Investigating: Toothpastes through Inquiry-Based Practical Work. Science Activities, 42(3), 31–37. https://doi.org/10.3200/SATS.42.3.31-37

Chi, S., Wang, Z., & Liu, X. (2019). Investigating disciplinary context effect on student scientific inquiry competence. International Journal of Science Education, 41(18), 2736–2764. https://doi.org/10.1080/09500693.2019.1697837

Clark, D., & Linn, M. C. (2003). Designing for knowledge integration: The impact of instructional time. The Journal of the Learning Sciences, 12(4), 451-493. https://doi.org/10.1207/S15327809JLS1204_1

Clarke-Midura, J., Code, J., Zap, N., & Dede, C. (2012). Assessing science inquiry: A case study of the Virtual Performance Assessment project. In L. Lennox, and K. Nettleton (Eds.). Cases on Inquiry through instructional technology in math and science: Systematic Approaches (pp. 138–164). IGI Global Publishing.

Clarke-Midura, J., Code, J., Zap, N., & Dede, C. (2012). Assessing science inquiry: A case study of the Virtual Performance Assessment project. In Lennex, L.C., & Nettleton, K.F. (Eds.). Cases on Inquiry through instructional technology in math and science (pp. 138-164). USA: IGI Global.

Crujeiras-Pérez, B., & Jiménez-Aleixandre, M. P. (2017). High school students' engagement in planning investigations: findings from a longitudinal study in Spain. Chemistry Education Research and Practice, 18(1), 99-112. https://doi.org/10.1039/C6RP00185H

Davies, D. J., Collier, C., & Howe, A. (2012). A matter of interpretation: developing primary pupils' enquiry skills using position-linked datalogging. Research in Science & Technological Education, 30(3), 311-325. https://doi.org/10.1080/02635143.2012.738420

Davies, D., Collier, C. & Howe, A. Assessing scientific and technological enquiry skills at age 11 using the e-scape system. International Journal of Technology and Design Education 22, 247–263 (2012). https://doi.org/10.1007/s10798-011-9191-3

Davis, J. P., & Bellocchi, A. (2018). Objectivity, subjectivity, and emotion in school science inquiry. Journal of Research in Science Teaching, 55(10), 1419–1447. https://doi.org/10.1002/tea.21461

Day, H. L., & Matthews, D. M. (2008). Do Large-Scale Exams Adequately Assess Inquiry? An Evaluation of the Alignment of the Inquiry Behaviors in New York State's Living Environment Regents Examination to the NYS Inquiry Standard. The American Biology Teacher, 70(6), 336-341. https://doi.org/10.1662/0002-7685(2008)70[336:DLEAAI]2.0.CO;2

De Jong, T. (2006). Technological advances in inquiry learning. Science, 312(5773), 532-533. https://doi.org/10.1126/science.112775

de Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. Science, 340(6130), 305–308. https://doi.org/10.1126/science.1230579

Demircioglu, T., & Ucar, S. (2015). Investigating the effect of argument-driven Inquiry in laboratory instruction. Educational Sciences: Theory & Practice, 15(1). https://doi.org/10.12738/estp.2015.1.2324

Di Mauro, M. F., & Furman, M. (2016). Impact of an inquiry unit on grade 4 students' science learning. International Journal of Science Education, 38(14), 2239-2258. https://doi.org/10.1080/09500693.2016.1234085

Donnelly, D. F., McGarr, O., & O'Reilly, J. (2014). 'Just Be Quiet and Listen to Exactly What He's Saying': Conceptualizing power relations in inquiry-oriented classrooms. International Journal of Science Education, 36(12), 2029-2054. https://doi.org/10.1080/09500693.2014.889867

Eslinger, E., White, B., Frederiksen, J., & Brobst, J. (2008). Supporting inquiry processes with an interactive learning environment: Inquiry Island. Journal of Science Education and Technology, 17(6), 610-617. https://doi.org/10.1007/s10956-008-9130-6

Fang, S. C., Hsu, Y. S., Chang, H. Y., Chang, W. H., Wu, H. K., & Chen, C. M. (2016). Investigating the effects of structured and guided Inquiry on students' development of conceptual knowledge and inquiry abilities: a case study in Taiwan. International Journal of Science Education, 38(12), 1945-1971. https://doi.org/10.1080/09500693.2016.1220688

Fang, S. C., Hsu, Y. S., Chang, H. Y., Chang, W. H., Wu, H. K., & Chen, C. M. (2016). Investigating the effects of structured and guided Inquiry on students' development of conceptual knowledge and inquiry abilities: a case study in Taiwan. International Journal of Science Education, 38(12), 1945-1971. https://doi.org/10.1080/09500693.2016.1220688

Garcia-Mila, M., Andersen, C., & Rojo, N. E. (2011). Elementary students' laboratory record keeping during Scientific Inquiry. International Journal of Science Education, 33(7), 915-942. https://doi.org/10.1080/09500693.2010.480986

Gobert, J. D., O'Dwyer, L., Horwitz, P., Buckley, B. C., Levy, S. T., & Wilensky, U. (2011). Examining the relationship between students' understanding of the nature of models and conceptual learning in biology, physics, and chemistry. International Journal of Science Education, 33(5), 653-684. https://doi.org/10.1080/09500691003720671

Haertel, G., Mislevy, R., Long, K., Kennedy, C., Wilson, M., & Songer, N. (2005). PADI: Principled assessment designs for Inquiry. Poster presented at the IERI PI Meeting, Washington, DC. Available at https://padi.sri.com/publications.html

Harlen, W. (2013). Assessment & Inquiry-Based Science Education: Issues in Policy and Practice. Global Network of Science Academies (IAP) Science Education Programme (SEP): Trieste, ItalyHaworth, C. M., Dale, P. S., & Plomin, R. (2010). Sex differences in school science performance from middle childhood to early adolescence. International journal of educational research, 49(2-3), 92-101. https://doi.org/10.1016/j.ijer.2010.09.003

Hein, G. E., & Lee, S. (2000). assessment of science inquiry. In National Science Foundation(Ed.), Foundations inquiry: Thoughts, views, and strategies for the k- 5 classroom (Vol. 2, pp.99-108). Arlington, VA: National Science Foundation. Available at https://nsf.gov/pubs/2000/nsf99148/htmstart.htm

Herranen, J., & Aksela, M. (2019). Student-question-based Inquiry in science education. Studies in Science Education, 55(1), 1–36. https://doi.org/10.1080/03057267.2019.1658059

Hickey, D. T., Ingram-Goble, A. A., & Jameson, E. M. (2009). Designing assessments and assessing designs in virtual educational environments. Journal of Science Education and Technology, 18(2), 187-208. https://doi.org/10.1007/s10956-008-9143-1

Hoffman, A., & Turner, K. (2015). Microbeads and engineering design in chemistry: No small educational investigation. Journal of Chemical Education, 92(4), 742-746. https://doi.org/10.1021/ed500623k

Hofstein, A., Shore, R., & Kipnis, M. (2004). Providing high school chemistry students with opportunities to develop learning skills in an inquiry-type laboratory: A case study. International Journal of Science Education, 26(1), 47-62. https://doi.org/10.1080/0950069032000070342

Hsu, C. C., Chiu, C. H., Lin, C. H., & Wang, T. I. (2015). Enhancing skill in constructing scientific explanations using a structured argumentation scaffold in scientific Inquiry. Computers & Education, 91, 46-59. https://doi.org/10.1016/j.compedu.2015.09.009

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(1), 1270-1280. https://doi.org/10.1016/j.compedu.2011.01.013

Ješková, Z., Lukáč, S., Hančová, M., Šnajder, Ľ., Guniš, J., Balogová, B., & Kireš, M. (2016). Efficacy of inquiry-based learning in mathematics, physics and informatics in relation to the development of students inquiry skills. Journal of Baltic Science Education, 15(5), 559.

Keselman, A. (2003). Supporting inquiry learning by promoting normative understanding of multivariable causality. Journal of Research in Science Teaching, 40, 898–921. https://doi.org/10.1002/Tea.10115.

Ketelhut, D. J., Nelson, B. C., Clarke, J., & Dede, C. (2010). A multi-user virtual environment for building and assessing higher-order inquiry skills in science. British Journal of Educational Technology, 41(1), 56–68. https://doi.org/10.1111/j.1467-8535.2009.01036.x

Klahr, D. (2000). Exploring science: The cognition and development of discovery processes. Cambridge: The MIT Press Kremer, K., Specht, C., Urhahne, D., & Mayer, J. (2014). The relationship in biology between the nature of science and scientific inquiry. Journal of Biological Education, 48(1), 1-8. https://doi.org/10.1080/00219266.2013.788541

Kruit, P., van den Berg, E., Schuitema, J., & Oostdam, R. (2017). Assessing inquiry skills of grade 5&6 students through performance assessments. In GIREP-ICPE-EPEC 2017: Bridging Research and Practice in Physics Teaching and Learning.

Kuang, X., Eysink, T. H. S., & Jong, T. (2020). Effects of providing partial hypotheses as a support for simulation-based inquiry learning. Journal of Computer Assisted Learning, 36(4), 487–501. https://doi.org/10.1111/jcal.12415

Kuo, C. Y., Wu, H. K., Jen, T. H., & Hsu, Y. S. (2015). Development and validation of a multimedia-based assessment of scientific inquiry abilities. International Journal of Science Education, 37(14), 2326-2357. https://doi.org/10.1080/09500693.2015.1078521

Kusmawan, U. (2024). Beyond Traditional Practicums: Exploring the Potential of Scalable Practicum in Science Courses. Studies in Learning and Teaching, 5(3), 622-637. https://doi.org/10.46627/silet.v5i3.505

Kusmawan, U. (2024). Transforming digital learning and assessment strategies in higher education. Multidisciplinary Reviews, 8(1), 2025016. https://doi.org/10.31893/multirev.2025016

Kyza, E. A., & Edelson, D. C. (2005). Scaffolding middle school students' coordination of theory and evidence. Educational Research and Evaluation, 11(6), 545-560. https://doi.org/10.1080/13803610500254857

Lau, K. C. (2017, December). Teaching about nature of science through short lab activities in Hong Kong classroom. Asia-Pacific Forum on Science Learning & Teaching, 18 (2).

Lederman, J., Lederman, N., Bartels, S., Jimenez, J., Akubo, M., Aly, S., Bao, C., Blanquet, E., Blonder, R., Soares de Andrade, B.S., Buntting, C., Cakir, M., Elzorkani, E.-D A., Gaigher, E., Guo, S., Hakanen, A., Al-lal, S.H., Han-Tosunoglu, C., Hattinhg, A., Hume, A., Irex, S. & Zhou, Q. (2019). An international collaborative investigation of beginning seventh-grade students' understandings of scientific Inquiry: Establishing a baseline. Journal of Research in Science Teaching, 56(4), 486-515. https://doi.org/10.1002/tea.21512

Lederman, N. G., Lederman, J. S., & Antink, A. (2013). Nature of science and scientific inquiry as contexts for the learning of science and achievement of scientific literacy. International Journal of Education in Mathematics, Science and Technology, 1(3).

Lee, J., & Yoon, J. Y. (2008). Teaching Early Childhood Teacher Candidates How to Assess Children's Inquiry Skills in Science Learning. Contemporary Issues in Early Childhood, 9(3), 265–269. https://doi.org/10.2304/ciec.2008.9.3.265

Leijen, Ä., Valtna, K., Leijen, D. A., & Pedaste, M. (2012). How to determine the quality of students' reflections?. Studies in Higher Education, 37(2), 203-217. https://doi.org/10.1080/03075079.2010.504814

Linn, M. C., & Eylon, B. S. (2006). Science education: Integrating views of learning and instruction. In P. A. Alexander& P. H. Winne (Eds.), Handbook of educational psychology (2nd ed., pp. 511–544). Mahwah, NJ: Erlbaum.

Linn, M. C., & Hsi, S. (2000). Computers, teachers, peers: Science learning partners. Mahwah, NJ: ErlbaumLinn, M. C., Davis, E. A., & Bell, P. (2004). Internet environments for science education. Mahwah, NJ: Erlbaum

Linn, M. C., & Hsi, S. (2000). Computers, teachers, peers: Science learning partners. Mahwah, NJ: Erlbaum

Liu, O. L., Lee, H. S., & Linn, M. C. (2010). Multifaceted assessment of inquiry-based science learning. Educational Assessment, 15(2), 69-86. https://doi.org/10.1080/10627197.2010.491067

Manlove, S., Lazonder, A. W., & de Jong, T. (2009). Trends and issues of regulative support use during inquiry learning: Patterns from three studies. Computers in Human Behavior, 25(4), 795-803. https://doi.org/10.1016/j.chb.2008.07.010

McNeill, K. L., Pimentel, D. S., & Strauss, E. G. (2013). The impact of high school science teachers' beliefs, curricular enactments and experience on student learning during an inquiry-based urban ecology curriculum. International Journal of Science Education, 35(15), 2608-2644. https://doi.org/10.1080/09500693.2011.618193

Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction–what is it and does it matter? Results from a research synthesis years 1984 to 2002. Journal of Research in Science Teaching, 47(4), 474–496. https://doi.org/10.1002/tea.20347

Mislevy, R. J., Chudowsky, N., Draney, K., Fried, R., Gaffney, T., & Haertel, G. (2003). Design Patterns for Assessing Science Inquiry. Menlo Park, CA: SRI International.

Moon, J. A., & Brockway, D. (2019). Facilitating learning in an interactive science simulation: The effects of task segmentation guidance on adults' inquiry-based learning and cognitive load. Journal of Research on Technology in Education, 51(1), 77–100. https://doi.org/10.1080/15391523.2019.1566038

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). TIMSS 2019 International Results in Mathematics and Science. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/international-results/

Mupira, P., & Ramnarain, U. (2018). The effect of inquiry-based learning on the achievement goal-orientation of grade 10 physical sciences learners at township schools in South Africa. Journal of Research in Science Teaching, 55(6), 810-825.

National Research Council. (1996). Introducing the National Science Education Standards. Washington, DC: The National Academy of Sciences. Retrieved from http://www.nap/edu/readingroom/books/intronses/

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

National Research Council. (2012). Next Generation Science Standards. Washington, DC: National Academy Press.

Nehring, A., Nowak, K. H., zu Belzen, A. U., & Tiemann, R. (2015). Predicting students' skills in the context of scientific Inquiry with cognitive, motivational, and sociodemographic variables. International Journal of Science Education, 37(9), 1343-1363. https://doi.org/10.1080/09500693.2015.1035358

Newman Jr, W. J., Abell, S. K., Hubbard, P. D., McDonald, J., Otaala, J., & Martini, M. (2004). Dilemmas of teaching inquiry in elementary science methods. Journal of Science teacher education, 15(4), 257-279. https://doi.org/10.1023/B:JSTE.0000048330.07586.d6

NGSS Lead States. (2013). Next Generation Science Standards: For states, by states. Washington: The National Academies Press.

Osborne, J. (2015). Practical work in science: Misunderstood and badly used. School Science Review, 96(357), 16–24.

Pedaste, M., Mäeots, M., Leijen, Ä., & Sarapuu, S. (2012). Improving students' inquiry skills through reflection and self-regulation scaffolds. Technology, Instruction, Cognition and Learning, 9(1–2), 81–95.

Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., Van Riesen, S. A., Kamp, E. T., … Tsourlidaki, E. (2015).

Peterson, S. M., & French, L. (2008). Supporting young children's explanations through inquiry science in preschool. Early childhood research quarterly, 23(3), 395-408. https://doi.org/10.1016/j.ecresq.2008.01.003

Phases of inquiry-based learning: Definitions and the inquiry cycle. Educational Research Review, 14, 47–61. https://doi.org/10.1016/ j.edurev.2015.02.003

Prayitno, B. A., Corebima, D., Susilo, H., Zubaidah, S., & Ramli, M. (2017). Closing the science process skills gap between students with high and low-level academic achievement. Journal of Baltic Science Education, 16(2), 266.

Quellmalz, E.S. & Pelligrino, J.W. (2009). Technology and testing. Science, 323,75–79.

Ramnarian (2014). Questioning the validity of inquiry assessment in a high stakes physical sciences examination. Perspectives in Education, 32 (1), 179 -191.

Ramnarian (2014). Teachers' perception of inquiry-based learning in urban, township and rural high schools: The context-specificity of science curriculum implementation in South Africa. Teaching and teacher education, 38, 65-75.

Rhea, M., Lucido, P., & Gregerson-Malm, C. (2005). Using Process and Inquiry to Teach Content: Projectile Motion and Graphing. Science Activities, 42(3), 10-15. https://doi.org/10.3200/SATS.42.3.10-15

Sampson, V., Enderle, P., Grooms, J., & Witte, S. (2013). Writing to learn by learning to write during the school science laboratory: Helping middle and high school students develop argumentative writing skills as they learn core ideas. Science Education, 97(5), 643-670. https://doi.org/10.1002/sce.21069

Scalise, K., Timms, M., Moorjani, A., Clark, L., Holtermann, K., & Irvin, P. S. (2011). Student learning in science simulations: design features that promote learning gains. Journal of Research in Science Teaching, 48(9), 1050–1078. https://doi.org/10.1002/tea.20437

Schwarz, C. V., & White, B. Y. (2005). Metamodeling knowledge: Developing students' understanding of scientific modelling. Cognition and instruction, 23(2), 165-205. https://doi.org/10.1207/s1532690xci2302_1

Shavelson, R. J., Baxter, G. P., and Pine, J. (1991). Performance assessments in science. Applied Measurement in Education, 4(4): 347-362.

Shivolo, T., & Mokiwa, H. O. (2024). Inquiry-Based Science Education: Perspectives from Namibian Teachers. International Journal of Research in STEM Education, 6(1), 97-112.

Shute, V. J., Hansen, E. G., & Almond, R. G. (2007). An assessment for learning system called ACED: Designing for learning effectiveness and accessibility. RR07-26. Princeton, NJ: Educational Testing Service.

Silk, E. M., Schunn, C. D., & Strand Cary, M. (2009). The impact of an engineering design curriculum on science reasoning in an urban setting. Journal of Science Education and Technology, 18(3), 209-223. https://doi.org/10.1007/s10956-009-9144-8

Songer, N. B., Lee, H. S., & McDonald, S. (2003). Research towards an expanded understanding of inquiry science beyond one idealized standard. Science Education, 87(4), 490-516. https://doi.org/10.1002/sce.10085

Stone, E. M. (2014). Guiding students to develop an understanding of scientific Inquiry: A science skills approach to instruction and assessment. CBE—Life Sciences Education, 13(1), 90-101. https://doi.org/10.1187/cbe-12-11-0198

Turkan, S., & Liu, O. L. (2012). Differential performance by English language learners on an inquiry-based science assessment. International Journal of Science Education, 34(15), 2343-2369. https://doi.org/10.1080/09500693.2012.705046

Van Uum, M. S., Verhoeff, R. P., & Peeters, M. (2016). Inquiry-based science education: towards a pedagogical framework for primary school teachers. International journal of science education, 38(3), 450-469. https://doi.org/10.1080/09500693.2016.1147660

Vitale, J. M., McBride, E., & Linn, M. C. (2016). Distinguishing complex ideas about climate change: knowledge integration vs specific guidance. International Journal of Science Education, 38(9), 1548-1569. https://doi.org/10.1080/09500693.2016.1198969

Wang, J., Guo, D., & Jou, M. (2015). A study on the effects of model-based inquiry pedagogy on students' inquiry skills in a virtual physics lab. Computers in Human Behavior, 49, 658-669. https://doi.org/10.1016/j.chb.2015.01.043

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modelling, and metacognition: Making science accessible to all students. Cognition and instruction, 16(1), 3-118. https://doi.org/10.1207/s1532690xci1601_2

Wu, H. K., Kuo, C. Y., Jen, T. H., & Hsu, Y. S. (2015). What makes an item more difficult? Effects of modality and type of visual information in a computer-based assessment of scientific inquiry abilities. Computers & Education, 85, 35-48. https://doi.org/10.1016/j.compedu.2015.01.007

Wu, P. H., Wu, H. K., & Hsu, Y. S. (2014). Establishing the criterion-related, construct, and content validities of a simulation-based assessment of inquiry abilities. International Journal of Science Education, 36(10), 1630-1650.https://doi.org/10.1080/09500693.2013.871660

Yeh, S. S. (2006). Tests worth teaching to: Constructing state-mandated tests that emphasize critical thinking. Educational Researcher, 30, 12–17.

Zuiker, S., & Whitaker, J. R. (2014). Refining inquiry with multi-form assessment: Formative and summative assessment functions for flexible Inquiry. International Journal of Science Education, 36(6), 1037-1059. https://doi.org/10.1080/09500693.2013.834489