Issues and Strategies in Inquiry-Based Learning Evaluation

Yiqing Mao

Faculty of Education, Beijing Normal University, Beijing, China
Email: maoyqstu@163.com

Abstract

Evaluation is an indispensable part of the inquiry-based learning. However, current research on the evaluation of inquiry-based learning still faces three problems: firstly, the evaluation system of inquiry-based learning is not clear enough; secondly, existing evaluation methods for inquiry-based learning often follow general learning evaluation standards, which do not reflect the true characteristics of inquiry-based learning, thus ignoring its specificity; and thirdly, evaluation of inquiry-based learning is prone to subjective influence from evaluators, and there is a lack of research on objective evaluation indicators. Scientific inquiry-based learning requires the establishment of an evaluation concept that can motivate and promote the coordinated development of students' knowledge and skills, processes and methods, emotional attitudes and values. This paper summarizes past research on evaluators, evaluation methods, and evaluation standards in inquiry-based learning, and proposes three strategies for future evaluation of inquiry-based learning.

Keywords

Inquiry-Based Learning, Learning Evaluation, Inquiry Process

1. Introduction

The concept of inquiry-based learning originated from Dewey's educational philosophy, emphasizing that scientific education should not just be about memorizing facts, but should also teach students how to think and act scientifically (Council, 2000). Inquiry-based learning is also known as "inquiry learning", "exploratory learning", "problem-based learning" and "discovery learning". Despite the diverse terminology, researchers (Jenkins, Healey, & Zetter, 2008; Kahn & O'Rourke, 2004; Weaver, 1989) generally agree on the core features of inquiry-based learning, namely: 1) Learning is inspired by inquiry, that is, driven

by questions or problems; 2) Learning is based on the process of constructing knowledge and generating new ideas; 3) It is an "active" learning method that includes learning through practice; 4) It is a student-centered teaching method, with teachers taking on the role of facilitators; 5) Students take increasing responsibility for their own learning.

For the past 50 years, inquiry-based learning has been recognized as an effective and popular method for providing scientific education, and in the last 20 years, it has gained widespread recognition in the educational community (McConney et al., 2014; Minner et al., 2010; Shymansky et al., 1990). Research shows that inquiry-based learning, based on the combination of scientific processes, knowledge, and reasoning, can enhance students' motivation to learn science (Crawford, 2014), deepen their understanding of concepts (Gott & Duggan, 2002), particularly in developing their self-reflection abilities, independent inquiry skills, critical thinking, and understanding of scientific concepts (Lederman et al., 2013; Lee, 2004; Saunders-Stewart et al., 2012), promote collaboration among students (Hofstein & Lunetta, 2004), enable students to use problem-solving research skills to create meaningful knowledge, thereby promoting deep learning (Buckner & Kim, 2014; Ellis & Bliuc, 2016), and has other unique advantages. Inquiry-based learning is usually conducted in groups, and during the learning process, students use evidence, logic, and imagination to understand the world around them (Newman Jr. et al., 2004).

Despite the repeated emphasis on the importance of inquiry-based learning in recent years, the quality of inquiry-based learning is a cause for concern (McConney, Oliver, Woods-McConney, Schibeci, & Maor, 2014). On the one hand, considering time constraints, crowded curricula, and learners' lack of experience with inquiry-based learning, introducing inquiry-based learning in basic science classes will pose a dilemma for teachers who lack sufficient experience with inquiry-based learning. Many teachers often fail to propose and guide students to explore thought-provoking inquiry questions and prefer to engage in Initiation-Response-Feedback (IRF) interactions, in which students only provide low-level thinking responses to teacher questions (Herbel-Eisenmann & Breyfogle, 2005). In a study of teacher-student interaction patterns in elementary school mathematics classrooms, scholars found that only about 15% of the questions posed by teachers were high-order questions or questions that required students to engage in critical thinking about the discussion topics (Wimer et al., 2001). In research on elementary school teacher questioning behavior, students are rarely asked cognitively challenging questions, which is precisely what activates critical thinking in students (Galton et al., 1999). This creates a dilemma: when teachers pose high-level questions that challenge students' thinking, peers provide more detailed help, which has a positive effect on their learning (Gillies, 2004; Gillies & Khan, 2008; Webb, 2009); however, when an IRF-like interaction occurs, this effect disappears—the role of the "teacher" becomes that of an actively problem-solving person, and the role of students becomes a "passive" person receiving guidance from the teacher, and they rarely share strategies for solving prob-

lems or explore their peers' ideas (Webb et al., 2008).

On the other hand, inquiry-based learning in the classroom has many challenging goals and unnecessary learning barriers that can cause students to lose focus. When engaging in inquiry-based learning, students must read, understand, and follow the teacher's guidance, collect data, interpret the data, and work collaboratively with their peers to communicate and exchange ideas. Therefore, the learning burden of inquiry-based learning is often heavy, and learners may suffer from information overload, making it difficult to sustain their focus on the expected learning goals. It is also common for students to become distracted by disorganized hands-on activities, fail to complete data collection within the designated time frame, make mistakes during observation, measurement, or recording due to improper use of instruments, lose interest in learning, or become too immersed in the details of their work to focus on the learning goals.

If the above issues are initially attributed to design flaws in inquiry-based learning, in order to optimize the design of inquiry-based learning, it is necessary to scientifically and reasonably evaluate its advantages and disadvantages and further improve its functionality. However, the current situation is that the evaluation of inquiry-based learning, whether in theory or practice, faces many challenges, such as deviation from learning objectives, overemphasis on inquiry results to the neglect of the inquiry process, lack of specific evaluation methods for inquiry-based learning, and so on. Without addressing the evaluation issue, it is impossible to provide optimization feedback for the design of inquiry-based learning, and there is great uncertainty in implementing inquiry-based learning to enhance the expected learning goals of students. Furthermore, the evaluation of inquiry-based learning is not a problem that can be solved solely through the teacher's experience; it requires systematic research.

2. Research on Inquiry-Based Learning

2.1. The Evaluators of Inquiry-Based Learning

The evaluators of inquiry-based learning are no different from those of general learning, mainly consisting of students and teachers. Furthermore, when students act as evaluators, the evaluation methods can be distinguished between peer evaluation and self-evaluation. Peer evaluations have three different types: intra-group, inter-group, and individual evaluations of the group (Earl, 1986). Intra-group peer evaluations are conducted when personal involvement of a student group working together is required, while inter-group evaluations may occur when a student group evaluates themselves or another group as a whole. Some studies suggest that participating in peer evaluations helps to improve the quality of learning outcomes, such as enhancing the quality of homework submitted by students after peer evaluations and revisions. Providing feedback to others is more conducive to learning than receiving feedback alone. These findings support peer evaluations becoming a driving force for learning tasks and a

powerful tool for self-evaluation.

Although peer evaluations have many benefits, there are still some unresolved issues, such as students feeling anxious about being evaluated by peers (J. Hanrahan; Liu & Carless, 2006), students not being honest enough in grading and providing critical feedback (Cho & Cho, 2011), limitations in student feedback leading to inaccurate evaluations (Dochy et al., 1999), and less benefit for high-achieving students in peer evaluations (Li & Gao, 2016; Ramon-Casas, Nuño, Pons, & Cunillera, 2019). Current solutions to these issues include conducting anonymous evaluations (Cho & Cho, 2011; Rotsaert, Panadero, & Schellens, 2018), requiring multiple students to evaluate the same work (Cho, Schunn, & Wilson, 2006), and using structured evaluation tools (Tsai & Chuang, 2013). However, these measures have two major problems. First, these solutions may bring new problems. Although anonymous peer evaluations reduce students' social emotional burden, they hinder student sharing, explaining ideas, and communicating during feedback activities (Ajjawi & Boud, 2017; Carless, 2012), and research has shown that anonymous evaluations reduce the accuracy of peer evaluations (Panadero & Alqassab, 2019). Having several students evaluate the same work can improve accuracy but may result in conflicting evaluation results, causing confusion for feedback recipients and decreasing learning enthusiasm (Wanner & Palmer, 2018). Not allowing students to grade deprives them of the opportunity to practice grading and evaluation skills. Second, these measures have not addressed the situation where "average" and "high-achieving" students benefit far less in peer evaluations than "low-achieving" students.

In order to avoid the inherent problems with using only teacher evaluations, combining peer and self-evaluations with teacher evaluations is a better approach. Teacher evaluations involve teachers assessing students, and they are important for understanding teaching quality and the impact of teachers on learners. However, when teacher evaluations are combined with self and peer evaluations, the traditional authoritative role of the teacher can be a "double-edged sword". While giving students the power to evaluate themselves and each other, the teacher's role can still influence students' judgments of their peers' performance. In other words, teachers may still dominate the evaluation process and be seen as the only standard in the classroom. Combining peer and self-evaluations helps teachers understand their students' learning situations and makes evaluations more comprehensive than either teacher or student evaluations alone. Through peer evaluations, teachers can understand most students' perspectives from grades, written comments, and oral feedback, which enhances teachers' understanding of high-quality behaviors displayed by students. Self-evaluation reflects each student's perception of their own performance. Therefore, the most feasible solution seems to be using self-evaluation to supplement peer evaluations to overcome the inherent problems of the latter.

Self-evaluation is a process by which students assess their own performance, also known as "self-feedback", "self-reflection", "self-review", and so on (Harris & Brown, 2018; Yan & Brown, 2017). Self-evaluation includes setting standards,

self-scoring, validating scores, and providing improvement suggestions (Boud, 1995; McMillan, 2013). Some scholars believe that self-evaluation includes accurately expressing expected performance, accurately expressing actual performance, taking action to close the gap between expected and actual performance (Sadler, 1989). Regardless of the interpretation, the content of self-evaluation can be summarized into three dimensions: goal awareness, performance awareness, and gap reduction.

Self-evaluation is closely related to reflection (Boud, 1995; Yan & Brown, 2017) because in the self-evaluation process, "the self is always available for personal use" (Harris & Brown, 2018). Reflection can be divided into reflection in action and reflection on action. The former is usually triggered unexpectedly during practice and if the evaluation process is not long, it will pass quickly, while the latter occurs after the event (Greenwood, 1993). Because it analyzes and interprets information from memory without time pressure, reflection on action is more systematic (Fitzgerald, 1994). Reflection in action is usually about optimizing on-site practice, while reflection on action is knowledge-oriented, allowing individuals to rethink the theories and concepts they identify with (Bolton, 2010). Although both are important, reflection on action involved in self-evaluation is a form of reflection that contributes to improving learning outcomes (Boud, 2013) and developing various skills, including self-regulation (Nicol & Macfarlane-Dick, 2006) and feedback literacy (Hoo, Deneen, & Boud, 2022; Tai, Ajjawi, Boud, Dawson, & Panadero, 2018; Yan & Carless, 2022). However, without explicit requirements from teachers, students mostly reflect in action rather than on action (Wanner & Palmer, 2018).

Like peer evaluation, self-evaluation also brings various benefits, including improving metacognition (Birjandi & Hadidi Tamjid, 2012), enhancing learning outcomes (Boud, 1995), and developing students' feedback literacy (Malecka, Boud, & Carless, 2022). If used alone, it also has its own problems. Some students resist this process (Boud, 1995), while others believe that they have already met the evaluation criteria (Hanrahan & Isaacs, 2001). This is called "self-blindness" by Rey (Rey, 2013). However, when combined with peer evaluation, self-evaluation can overcome most of the problems (Dochy, Segers, & Sluijsmans, 1999; Hung, Samuelson, & Chen, 2016; Topping, 2018). Self-evaluation is a supplement to peer evaluation, and when peer evaluation is lacking, self-evaluation enables students to engage in self-reflection. Even when students have the opportunity to engage in peer evaluation, self-evaluation can effectively supplement peer evaluation because these two processes involve reflection from different perspectives.

2.2. Evaluation Methods for Inquiry-Based Learning

This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire journals, and not as an independent document. Please do not revise any of the current designations. The methods for evaluation mainly rely on tests, portfolio, rating scales, and so on. In recent years, in-

formation technology and learning management systems have also been added to collect process data in the field of learning analytics. For example, Wu et al. (2022) developed a process data-based collaborative problem-solving automation evaluation system relying on the PSAA platform, which can automatically collect trigger behavior, process behavior data (clickstream data and session data), and so on. In terms of using portfolio to collect process data, Herman (1992) used portfolio evaluation as the basis for teacher and student evaluation. A portfolio is defined as "a collection of student work reviewed according to standards to evaluate a student or a project." Types of student work can include essays, videos, etc. To use a portfolio to evaluate students, teachers should clearly define the evaluation purpose, determine the method for determining the skills that should be included in the portfolio, such as problem-solving skills and communication skills, who and when should determine this, and determine the sample selection and judgment criteria.

In terms of using rating scales to collect process data, Huang and Ma (2012) developed an observable and interpretable evaluation standard rating scale, which includes four observation dimensions: exploratory subject, exploratory interaction, exploratory resources, and exploratory ability. Each dimension has 4 - 11 different observation items. In the exploratory subject, the focus is on who the classroom subject is and the overall change of the classroom subject. In exploratory interaction, the focus is on the interaction between different teachers and students and different teaching links. In exploratory resources, the focus is on the development of resources in different schools, teaching processes, teaching links, and course types. In exploratory ability, the focus is on cultivating the ability to propose questions, cognitive and interpretive abilities, production and design abilities, communication and expression abilities, and so on.

We believe that evaluating the quality of exploratory learning only from the changes in the exploratory subject in the classroom, teacher-student interactions in exploratory interactions, exploratory resources, and exploratory ability may not be comprehensive. Moreover, how to measure students' abilities to propose questions, communicate and express themselves is a tricky problem. This may be evaluated by students themselves, peers, or teachers. However, questionnaire-based evaluation is a subjective process and is an impression of the overall course, which is not scientific and realistic. Teachers and students may ignore some details, which precisely reflect the performance of exploration or higher-order skills formed through exploration.

This passage discusses two basic orientations in learning evaluation, process-oriented and outcome-oriented, which are differentiated based on the nature of the detail data, whether they are process data or result data (Glasgow, 1997; Swanson, Case, & van der Vleuten, 1991). Process-oriented evaluation emphasizes students' ability to solve problems using knowledge, including recording observations of student learning logs, measuring communication skills in class, using learning resources, and developing problem-solving skills. Outcome-oriented evaluation indirectly measures the quality of the learning process by examining

the results after a period of learning, believing that the learning products created by students reflect their ability to apply knowledge (Swanson et al., 1991). This can include grading student work using scoring criteria, or evaluating based on standardized tests or surveys.

In the outcome-oriented evaluation orientation, Li et al. (2022) proposed an Outcome-Oriented Pattern-Based Model (OOPB), which is a result-oriented evaluation model that begins with clearly defining the learning outcomes and then reverse designs the learning tasks and evaluation to promote the expected learning outcomes. Teachers need to strengthen systematic course design methods and prioritize consistency between learning outcomes, tasks, and evaluation (Biggs, 1996; Biggs, 2014; Biggs & Tang, 2011). In an empirical study by Kogan and Laursen (2014), a standardized method was developed to calculate and average students' grades after taking inquiry-based learning courses and non-inquiry-based learning courses, and it was found that students who took inquiry-based learning courses performed better on their other courses. This is a way to evaluate the quality of inquiry-based learning by representing it through students' academic performance. However, many people question whether current standardized tests match important goals for student learning and development, as they only measure basic skills in reading, language, and math, and sometimes do not match the curriculum and teaching, ignoring the measurement of complex thinking processes and problem-solving skills (Herman, 1992).

There are three types of process-oriented evaluations, the first being self-evaluation tools, which include self-evaluation scales. Researchers have developed detailed and rich scales using nested designs that iterate around factors, dimensions, sub-scales, sub-dimensions, and specific items. Li and Wang (2021) designed a self-assessment scale for college students' extracurricular autonomy based on six evaluation dimensions and 15 sub-questions, including "Do you pay attention to extending the content of classroom learning?" and "Do you pay attention to expanding your learning interests?" Using this scale to evaluate students' learning processes can improve the accuracy of their answers due to its high level of precision and strong guidance.

The second type is structured interview tools (Zimmerman & Pons, 1986), which involve face-to-face conversations that prompt students to recall their learning behaviors. This method is more open-ended and avoids the risk of constraining learners' responses through a predefined questionnaire. Learners can add their own thoughts and answers based on the structured interview tool's questions. This evaluation method is often used in combination with the first type.

The third type is video analysis tools, which can capture all the behaviors of students and teachers in the classroom (Fischer & Neumann, 2012). For example, Schreiber et al. (2016) recorded videos of 14 students conducting physics experiments to evaluate their skills. Trained coders cut the videos into 5-second segments and categorized the students' behaviors, such as designing experimental plans, operating instruments, processing data, explaining results, and filling out experiment sheets. Then they evaluated whether each category was per-

formed correctly, imperfectly, or incorrectly.

Overall, although evaluation tools such as scales or structured interviews have strong operational feasibility, there is a possibility of learners' subjective bias. That is, students' responses may not be based on actual situations. In addition, the objectivity of evaluation results from self-evaluation tools is not optimistic, even though the learners in high school and above have no difficulty understanding the contents of the evaluation tools (Sun & Zheng, 2021). Moreover, for learners in junior high school and even primary school, understanding the content dimensions of evaluation tools can be difficult. For example, there is considerable uncertainty in how to distinguish between learning strategies and learning outcomes, and how to understand learning objectives. It is also unclear whether this really evaluates process data, and it may be better understood as phased result data.

The number of papers that truly evaluate process data is relatively scarce in academic writing. The instructional process mechanism diagram designed by Yang et al. (2017) covers learning objectives and content, as well as participants and their behaviors, and the process of transforming teaching content themes, providing a detailed expression of the interactive processes that teachers and students engage in regarding specific topics. In addition, Yang K et al. have also developed a role behavior coding system for collaborative learning activities, which allows for the characterization of specific behaviors and interactions between teachers and students using this system (Yang et al., 2017). Furthermore, Wu et al. (2022) have developed an automated evaluation framework for collaborative problem-solving abilities based on process data flow, which automatically extracts evidence for evaluation from process behavior data (clickstream and session data) generated by individuals in the assessment system to evaluate students' collaborative problem-solving abilities.

There is no one correct way to evaluate students. Although outcome-oriented evaluations can be persuasive, we do not advocate for this type of evaluation for all types of learning, nor do we reject process-oriented evaluations for specific learning processes. What is worthy of recognition is that outcome-oriented evaluations provide a powerful method for evaluating complex thinking and problem-solving abilities, and because they are based on real-world interests, they may be more motivating and reinforcing for students. However, although outcome-oriented evaluations can reflect the degree and depth to which students apply knowledge, process-oriented evaluations may be more effective in determining the extent to which students have mastered basic facts and concepts, as well as in reflecting students' high-level abilities. In summary, we need to determine the most appropriate evaluation methods based on the specific nature of the learning process.

2.3. Assessment Criteria for Inquiry-Based Learning

The standards used to evaluate student performance are referred to as grading criteria or grading guidelines. In order to make a value judgment about an object

of value, one must not only have access to process data related to the object but must also establish evaluation criteria. The process of evaluating an object of value is essentially a process of making value judgments. Such judgments must be based on a standard or scale, which is the basis for making judgments. This standard may be based on a crude, unipolar indicator system such as ideology, values, or a slightly more complex multi-level indicator system.

For example, Ruan and Zhang (2014) proposed that evaluation of inquiry-based learning should be based on the following principles: 1) the problem situation must be directional, inspiring, and closely related to reality, with exploratory value; 2) knowledge transmission and exploration processes should be combined organically to make classroom teaching more efficient; 3) starting from the exploration process, inquiry-based learning not only requires the acquisition of knowledge, but also aims to exercise inquiry skills and the ability to use inquiry methods to solve practical problems; 4) a harmonious teacher-student relationship should be fostered in inquiry-based teaching activities, and students should be encouraged to question and speculate, thus stimulating their potential for inquiry.

Researchers have also developed more complex multi-level indicator systems to establish criteria for evaluating individual indicators. Liu et al. (2022) proposed a three-level evaluation indicator system for inquiry-based learning based on the CIPP evaluation model framework, and used the Delphi method for revision to address issues such as the lack of improvement mechanisms in the evaluation of the learning process. Lu et al. (2013) designed a performance evaluation form for student inquiry-based learning courses based on learning attitudes, co-operative spirit, and inquiry processes, an evaluation form for student works based on three aspects: ideology and science, creativity, and artistry, and a teacher's teaching performance evaluation form based on teaching content, teaching methods, and teaching effects. For example, in the evaluation of the inquiry process, the main evaluation indicators are: 1) the scientific, feasible, novel, and practical selection of topics; 2) the reliability of experimental or investigative data; 3) the appropriateness of research methods; 4) strong practical skills; 5) innovation consciousness and skills. The evaluation criteria are divided into four levels: "excellent", "good", "average", and "poor" with scores ranging from 4 to 1. However, the evaluation criteria based on an indicator system cannot guarantee that the evaluations made are accurate.

3. Shortcomings in Evaluating Inquiry-Based Learning

At present, there is still a relative scarcity and preliminary nature of research specifically aimed at exploring learning evaluation, with most existing studies adopting a general approach to handling evaluation in inquiry-based learning. As a result, the unique characteristics of inquiry-based learning are difficult to reflect in its evaluation.

Inquiry-based learning aims to change students' learning styles, highlighting their cognitive activities such as exploration and discovery during the learning

process, making learning more about the process of students identifying and solving problems (Huang & Ma, 2012). The unique features of inquiry-based learning, such as student-initiated learning, learning driven by inquiry, problem-based or problem-driven learning, and learning based on knowledge construction and generating new ideas, require that the dimensions and framework for evaluating inquiry-based learning should not be equated with traditional teaching. However, research on evaluating inquiry-based learning, apart from emphasizing the basic principles that evaluation should follow (such as holistic and open principles) (Huang & Ma, 2012), is mainly aimed at setting up a set of evaluation indicators. For example, some studies believe that the evaluation index system for physics inquiry-based classroom teaching should mainly be based on the following aspects: 1) teaching objectives, 2) teaching content, 3) teaching process, and 4) process rationality (Huang & Ma, 2012).

However, the indicators in these types of systems are too abstract and general, and the scoring of each indicator is mainly based on subjective feelings, which can easily lead to bias. In a short period of time, neither experts nor teachers can score dozens of indicators based solely on watching the implementation process of inquiry-based learning. In addition, the dimensions corresponding to the truth of the value standards and value objects must be consistent. Evaluation should be based on process data of the evaluation object. If a scoring system is used, the expert's scoring is not based on the truth, but on their feelings and experience of the entire process. This is not an evaluation of the inquiry-based learning process.

The aim of inquiry-based learning is to enable students to utilize problem-solving skills to create meaningful knowledge, promoting deep learning (Buckner & Kim, 2014; Ellis & Bliuc, 2016) and developing critical and higher-order thinking skills (Bush, Sieber, Seiler, & Chandler, 2017). This means that the curriculum must not only include scientific concepts and methods, but also achieve knowledge and skill goals, and demonstrate inquiry behavior. Looking back at research on inquiry-based learning evaluation, it is difficult to find literature that reflects specific student inquiry behavior and evaluation indicators closely related to pre-set learning goals. Even if a few evaluation indicators reflect student inquiry behavior, it is difficult to reflect the higher-order abilities acquired by specific inquiry-based learning. Perhaps the reason why the evaluation system of inquiry-based learning is not clear and its discussion is relatively difficult is that our existing methods of learning evaluation are not suitable for inquiry-based learning.

The inquiry-based learning we are studying is inquiry-based learning in real educational contexts, and therefore "inquiry" and "achieving specific learning goals set by the educational context" are its fundamental characteristics. If evaluation is conducted outside of these two points, it is no different from general learning evaluation. As mentioned earlier, learning evaluation can be divided into two orientations: process-oriented and result-oriented. Due to the uniqueness of inquiry-based learning itself, it is not appropriate to use a result-oriented approach for its evaluation, only a process-oriented approach can be adopted. Learning outcomes that students produce over a period of time, such as daily

quizzes, homework assignments, and learning works, are difficult to demonstrate whether they have achieved the high-order ability learning goals required by inquiry-based learning because the results are obviously subject to the student's personal efforts. From this perspective, even if students do not produce good learning products within the limited time of the classroom, it does not mean that they have not completed good inquiry-based learning. Moreover, a result-oriented evaluation method is not based on the truth of the learning process, but on the truth of the learning outcomes, which seems to imply that mastering the truth of the learning outcomes is equivalent to mastering the truth of the learning process. This is not a value-based reasoning but a causal logical reasoning, and it has flaws. We cannot infer good processes from good results.

However, challenges exist in adopting a process-oriented approach in evaluating inquiry-based learning. Firstly, the collection of process data for evaluation requires skill, as it is complex to evaluate data generated in the early stages of inquiry. Since students' inquiry paths may change, the final inquiry path they adopt may not be the same as the initial one (Biggers, Forbes, & Zangori, 2013). If data is collected too early, key ideas or misconceptions of students may be missed, however, if data is collected too late, they may miss the momentary performance of students' inquiry. Therefore, it is difficult for teachers to determine which type of evaluation is most effective and when to evaluate during class. In addition, the evaluation ability of teachers during specific inquiry processes is continuously improving, and sometimes it is realized in the later stages of inquiry process that selecting the best time for evaluation is crucial, which may affect the reliability of process data. Secondly, evaluation of inquiry-based learning is subject to subjective interference and cannot reflect specific inquiry behaviors or learning objectives. Although the assessment indicators and evaluation criteria for each stage are based on process data, unfortunately, the indicators are often general, such as attitudes, approaches, and qualities in problem solving (Wang, He, & Zhang, 2010), which are not enough to reflect specific inquiry behaviors or learning objectives. Worse still, evaluators often score based on subjective feelings rather than objective facts, which can easily lead to errors and differences between evaluators.

The evaluation of inquiry-based learning is in a dilemma, as evaluating inquiry learning based on its outcome is not entirely reasonable, and evaluating inquiry learning process with a set of indicators is not convincing enough. The approach and methods of evaluating inquiry-based learning need further improvement by researchers.

4. Suggestions for Future Evaluation of Inquiry-Based Learning

4.1. The Principle of Inquiry-Based Learning Evaluation Should Be Clearly Defined

To begin with, it is necessary to clarify the principles of evaluation for inquiry-

based learning.

1) Learning behaviors should correspond to the target knowledge points. The target knowledge points refer to the knowledge content that students acquire through inquiry-based learning, such as the functional indicators of a bridge, which include stability, load-bearing capacity, and trafficability, in a bridge design course. The various inquiry-based learning stages of the course are conducted in a certain order, and the knowledge points contained in each stage are different. Each of the student's behaviors, such as experimental operations, proposing new ideas or questioning, must correspond to the knowledge points that the student has obtained or already knows. If a student's inquiry behavior in this stage cannot reflect the relevant knowledge points, it indicates that the efficiency of this stage is not high, and a lower score should be given in the evaluation.

2) Learning behaviors should reflect the characteristics of inquiry forms. Some behaviors not only exist in the inquiry classroom but also commonly exist in general learning, which we call irrelevant behaviors. For example, questioning about operation methods or design ideas during the discussion process, or asking for help, all of which cannot reflect the formative characteristics of inquiry. For irrelevant behaviors that occur in the inquiry classroom, the more frequent they occur, the lower the score should be given in the evaluation. However, behaviors such as "questioning", "discussing", "inquiring" and "operating" that reflect the form and characteristics of inquiry-based learning should be given higher scores in the evaluation when they occur more frequently.

3) Learning behaviors should reflect specific high-level abilities acquired in inquiry-based learning. In the process of inquiry-based learning, students will acquire certain high-level abilities, such as innovative spirit, scientific literacy, computational thinking, critical thinking, and so on. Therefore, in the evaluation, higher scores should be given to these factors. Learning behaviors should reflect the specific high-level abilities acquired in inquiry-based learning. Students will acquire some high-level abilities in inquiry-based learning, such as innovative spirit, scientific literacy, computational thinking, and critical thinking. When designing the evaluation system and evaluation indicators, the above dimensions should be taken into account.

4.2. The Evaluation of Inquiry-Based Learning Should Mainly Focus on Process Evaluation

The purpose of evaluating inquiry-based learning is to provide better feedback. If only results-oriented evaluation is emphasized, it is difficult to identify problems that exist throughout the entire inquiry process. Moreover, inquiry-based learning itself is a process in which students construct scientific knowledge, generate and enhance scientific attitudes, master scientific methods, and promote creative problem-solving abilities autonomously through a repeated cycle of "perceiving problems-defining problems-solving problems" in a real educational context. Therefore, process data should be recorded in the evaluation to provide more targeted feedback. An actionable method is to record the process of in-

quiry-based learning, and then cut the behavior data of teachers and students into sentences and items in video analysis. The behavior data is encoded in the order of "behavior subject-specific behavior-sequence number-knowledge point sequence-interaction", which can reflect the dynamic participation process of students and teachers in the inquiry-based classroom and also the interaction process among different members of the group. For example, we use S_i to represent the student sequence, K_i to represent the knowledge sequence, and T to represent the teacher. The behavior number is coded in three-digit sequence starting from 000. The interaction between teachers and students or among students is marked with the [ ], and the subsequent action caused by this action is described within the [ ]. S1-BM-010 {K1, K2, K6, K9} [→S2-TL-020] means that student S1 expressed his or her opinion, the current behavior number of S1 is 010, the knowledge points contained in the opinion include K1, K2, K6, and K9, and S1's behavior at this time triggered a discussion expressed by S2 with a behavior number of 020.

4.3. Compare and Evaluate Teaching Design with Inquiry Process

Once the dynamic process of joint participation of students and teachers in the exploration course and the process of interaction among different members of the group are obtained, it is necessary to develop evaluation criteria to assess these behavioral data, which should correspond to the original instructional design. That is to say, for the exploration performance and path in the instructional design, evaluation scores should be given based on how well they are achieved, and scores should also be given for any missing or deviated aspects. First, data such as target knowledge points, exploration tasks, exploration paths, key interactions between teachers and students and among students, and key behaviors should be extracted separately from the instructional design and the actual exploration process, which can indirectly reflect students' exploration process. Then, the data extracted from the instructional design and the actual exploration process should be compared and classified and coded according to certain operational steps. This can demonstrate what opportunities for improving exploratory learning are hidden in the differences between the actual exploration process and the instructional design. The advantage of this approach is that evaluation only needs to compare the relevant data between the actual exploration process and the instructional design, and the value interpretation can be made through the evaluation criteria. This avoids subjective errors, simplifies the evaluation process, and even without the involvement of experts, evaluation of exploratory learning can still be made.

5. Conclusion

Inquiry-based learning, a student-centered and problem-solving approach, has become increasingly popular in the classroom. However, this teaching method is not without its challenges. Firstly, time management and planning can be a significant issue. As this learning method requires more time for exploration and

problem-solving, it is essential to plan time and resources carefully to ensure students can complete tasks within the allocated time. Secondly, student autonomy can be problematic as inquiry-based learning requires students to have self-management and learning skills, but some students may face difficulties with self-motivation and management. Thirdly, the teacher's role in inquiry-based learning is to guide and support students rather than directly imparting knowledge. This requires teachers to have different skills and approaches to support student learning, including how to pose questions, guide student thinking, encourage exploration, and provide feedback. Lastly, inquiry-based learning requires a different assessment method than traditional teaching methods. Teachers need to understand students' learning progress and outcomes and provide real-time feedback and support. Evaluation methods for inquiry-based learning should mainly focus on process evaluation, comparing teaching design with inquiry process evaluation. At the same time, the following three principles should be followed: learning behavior should correspond to the target knowledge points, reflect the formal characteristics of the inquiry process, and reflect the high-order abilities acquired in specific inquiry-based learning.

In conclusion, inquiry-based learning poses several challenges in the classroom, such as time management, student autonomy, teacher roles, and assessment feedback. However, with proper solutions, these challenges can be overcome, and the benefits of inquiry-based learning, such as promoting student creativity, critical thinking, and problem-solving skills, can be realized.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Ajjawi, R., & Boud, D. (2017). Researching Feedback Dialogue: An Interactional Analysis Approach. *Assessment & Evaluation in Higher Education, 42,* 252-265. https://doi.org/10.1080/02602938.2015.1102863

Biggers, M., Forbes, C. T., & Zangori, L. (2013). Elementary Teachers' Curriculum Design and Pedagogical Reasoning for Supporting Students' Comparison and Evaluation of Evidence-Based Explanations. *The Elementary School Journal, 114,* 48-72. https://doi.org/10.1086/670738

Biggs, J. (1996). Enhancing Teaching through Constructive Alignment. *Higher Education, 32,* 347-364. https://doi.org/10.1007/BF00138871

Biggs, J. (2014). Constructive Alignment in University Teaching. *HERDSA Review of Higher Education, 1,* 5-22.

Biggs, J., & Tang, C. (2011). *Teaching for Quality Learning at University.* McGraw-Hill Education (UK).

Birjandi, P., & Hadidi Tamjid, N. (2012). The Role of Self-, Peer and Teacher Assessment in Promoting Iranian EFL Learners' Writing Performance. *Assessment & Evaluation in Higher Education, 37,* 513-533. https://doi.org/10.1080/02602938.2010.549204

Bolton, G. (2010). *Reflective Practice: Writing and Professional Development.* Sage Pub-

lications.

Boud, D. (1995). *Enhancing Learning through Self-Assessment.* Routledge Falmer.

Boud, D. (2013). Self and Peer Marking in a Large Technical Subject. In *Enhancing Learning through Self-Assessment* (pp. 71-86). Routledge. https://doi.org/10.4324/9781315041520-11

Buckner, E., & Kim, P. (2014). Integrating Technology and Pedagogy for Inquiry-Based Learning: The Stanford Mobile Inquiry-Based Learning Environment (SMILE). *Prospects, 44,* 99-118. https://doi.org/10.1007/s11125-013-9269-7

Bush, D., Sieber, R., Seiler, G., & Chandler, M. (2017). University-Level Teaching of Anthropogenic Global Climate Change (AGCC) via Student Inquiry. *Studies in Science Education, 53,* 113-136. https://doi.org/10.1080/03057267.2017.1319632

Carless, D. (2012). Trust and Its Role in Facilitating Dialogic Feedback. In *Feedback in Higher and Professional Education* (pp. 90-103). Routledge.

Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and Reliability of Scaffolded Peer Assessment of Writing from Instructor and Student Perspectives. *Journal of Educational Psychology, 98,* 891. https://doi.org/10.1037/0022-0663.98.4.891

Cho, Y. H., & Cho, K. (2011). Peer Reviewers Learn from Giving Comments. *Instructional Science, 39,* 629-643. https://doi.org/10.1007/s11251-010-9146-1

Council, N. R. (2000). *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning.* National Academies Press.

Crawford, B. A. (2014). From Inquiry to Scientific Practices in the Science Classroom. In *Handbook of Research on Science Education, Volume II* (pp. 529-556). Routledge. https://doi.org/10.4324/9780203097267-36

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The Use of Self-, Peer and Co-Assessment in Higher Education: A Review. *Studies in Higher Education, 24,* 331-350. https://doi.org/10.1080/03075079912331379935

Earl, S. E. (1986). Staff and Peer Assessment-Measuring an Individual's Contribution to Group Performance. *Assessment and Evaluation in Higher Education, 11,* 60-69. https://doi.org/10.1080/0260293860110105

Ellis, R. A., & Bliuc, A. M. (2016). An Exploration into First-Year University Students' Approaches to Inquiry and Online Learning Technologies in Blended Environments. *British Journal of Educational Technology, 47,* 970-980. https://doi.org/10.1111/bjet.12385

Fischer, H. E., & Neumann, K. (2012). Video Analysis as a Tool for Understanding Science Instruction. In *Science Education Research and Practice in Europe* (pp. 115-139). Brill. https://doi.org/10.1007/978-94-6091-900-8_6

Fitzgerald, M. (1994). *Theories of Reflection for Learning.*

Galton, M., Hargreaves, L., Comber, C., Wall, D., & Pell, T. (1999). Changes in Patterns of Teacher Interaction in Primary Classrooms: 1976-96. *British Educational Research Journal, 25,* 23-37. https://doi.org/10.1080/0141192990250103

Gillies, R. M. (2004). The Effects of Communication Training on Teachers' and Students' Verbal Behaviours during Cooperative Learning. *International Journal of Educational Research, 41,* 257-279. https://doi.org/10.1016/j.ijer.2005.07.004

Gillies, R. M., & Khan, A. (2008). The Effects of Teacher Discourse on Students' Discourse, Problem-Solving and Reasoning during Cooperative Learning. *International Journal of Educational Research, 47,* 323-340. https://doi.org/10.1016/j.ijer.2008.06.001

Glasgow, N. A. (1997). *New Curriculum for New Times: A Guide to Student-Centered, Problem-Based Learning.* ERIC.

Gott, R., & Duggan, S. (2002). Problems with the Assessment of Performance in Practical Science: Which Way Now? *Cambridge Journal of Education, 32,* 183-201. https://doi.org/10.1080/03057640220147540

Greenwood, J. (1993). Reflective Practice: A Critique of the Work of Argyris and Schön. *Journal of Advanced Nursing, 18,* 1183-1187. https://doi.org/10.1046/j.1365-2648.1993.18081183.x

Hanrahan, J. S., & Isaacs, G. (2001). Assessing Self- and Peer-Assessment: The Students' Views. *Higher Education Research & Development, 20,* 53-70. https://doi.org/10.1080/07294360123776

Harris, L. R., & Brown, G. T. (2018). *Using Self-Assessment to Improve Student Learning.* Routledge. https://doi.org/10.4324/9781351036979

Herbel-Eisenmann, B. A., & Breyfogle, M. L. (2005). Questioning Our Patterns of Questioning. *Mathematics Teaching in the Middle School, 10,* 484-489. https://doi.org/10.5951/MTMS.10.9.0484

Herman, J. L. (1992). *A Practical Guide to Alternative Assessment.* ERIC.

Hofstein, A., & Lunetta, V. N. (2004). The Laboratory in Science Education: Foundations for the Twenty-First Century. *Science Education, 88,* 28-54. https://doi.org/10.1002/sce.10106

Hoo, H.-T., Deneen, C., & Boud, D. (2022). Developing Student Feedback Literacy through Self and Peer Assessment Interventions. *Assessment* & *Evaluation in Higher Education, 47,* 444-457. https://doi.org/10.1080/02602938.2021.1925871

Huang, H., & Ma, Y. (2012). An Exploration of the Implementation of Inquiry-Based Teaching Evaluation: SIRA Classroom Observation Mode. *Education Development Research, 32,* 63-66.

Hung, Y.-J., Samuelson, B. L., & Chen, S.-C. (2016). Relationships between Peer- and Self-Assessment and Teacher Assessment of Young EFL Learners' Oral Presentations. In *Assessing Young Learners of English: Global and Local Perspectives* (pp. 317-338). Springer. https://doi.org/10.1007/978-3-319-22422-0_13

Jenkins, A., Healey, M., & Zetter, R. (2008). *Linking Teaching and Research in Disciplines and Departments.*

Kahn, P., & O'Rourke, K. (2004). Guide to Curriculum Design: Enquiry-Based Learning. *Higher Education Academy, 30,* 3-30.

Kogan, M., & Laursen, S. L. (2014). Assessing Long-Term Effects of Inquiry-Based Learning: A Case Study from College Mathematics. *Innovative Higher Education, 39,* 183-199. https://doi.org/10.1007/s10755-013-9269-9

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,* 138-147.

Lee, V. S. (2004). *Teaching and Learning through Inquiry: A Guidebook for Institutions and Instructors.* Stylus Pub LLC.

Li, D., & Wang, Y. (2021). Assessment and Analysis of Extracurricular Autonomous Learning Ability of College Students: A Case Study of Beihang University. *Journal of Beihang University (Social Sciences Edition), 34,* 111-117.

Li, L., & Gao, F. (2016). The Effect of Peer Assessment on Project Performance of Students at Different Learning Levels. *Assessment* & *Evaluation in Higher Education, 41,* 885-900. https://doi.org/10.1080/02602938.2015.1048185

Li, L., Farias Herrera, L., Liang, L., & Law, N. (2022). An Outcome-Oriented Pattern-Based Model to Support Teaching as a Design Science. *Instructional Science, 50,*

111-142. https://doi.org/10.1007/s11251-021-09563-4

Liu, N.-F., & Carless, D. (2006). Peer Feedback: The Learning Element of Peer Assessment. *Teaching in Higher Education, 11,* 279-290.
https://doi.org/10.1080/13562510600680582

Liu, W., Chen, Y., Liu, S., Wang, Y., & Huang, X. (2022). Construction of Evaluation Index System for Inquiry-Based Teaching. *Journal of Liaoning University of Technology (Social Science Edition), 24,* 116-120.

Lu, C., Jiang, R., & Deng, Q. (2013). Construction of Evaluation Index System for Inquiry-Based Teaching Curriculum. *University Teaching in China, No. 6,* 76-78+88.

Malecka, B., Boud, D., & Carless, D. (2022). Eliciting, Processing and Enacting Feedback: Mechanisms for Embedding Student Feedback Literacy within the Curriculum. *Teaching in Higher Education, 27,* 908-922. https://doi.org/10.1080/13562517.2020.1754784

McConney, A., Oliver, M. C., Woods-Mcconney, A., Schibeci, R., & Maor, D. (2014). Inquiry, Engagement, and Literacy in Science: A Retrospective, Cross-National Analysis Using PISA 2006. *Science Education, 98,* 963-980. https://doi.org/10.1002/sce.21135

McMillan, J. H. (2013). *SAGE Handbook of Research on Classroom Assessment.* Sage.
https://doi.org/10.4135/9781452218649

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: The Official Journal of the National Association for Research in Science Teaching, 47,* 474-496.
https://doi.org/10.1002/tea.20347

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,* 257-279.
https://doi.org/10.1023/B:JSTE.0000048330.07586.d6

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative Assessment and Self-Regulated Learning: A Model and Seven Principles of Good Feedback Practice. *Studies in Higher Education, 31,* 199-218. https://doi.org/10.1080/03075070600572090

Panadero, E., & Alqassab, M. (2019). An Empirical Review of Anonymity Effects in Peer Assessment, Peer Feedback, Peer Review, Peer Evaluation and Peer Grading. *Assessment & Evaluation in Higher Education, 44,* 1253-1278.
https://doi.org/10.1080/02602938.2019.1600186

Ramon-Casas, M., Nuño, N., Pons, F., & Cunillera, T. (2019). The Different Impact of a Structured Peer-Assessment Task in Relation to University Undergraduates' Initial Writing Skills. *Assessment & Evaluation in Higher Education, 44,* 653-663.
https://doi.org/10.1080/02602938.2018.1525337

Rey, G. (2013). We Are Not All "Self-Blind": A Defense of a Modest Introspectionism. *Mind & Language, 28,* 259-285. https://doi.org/10.1111/mila.12018

Rotsaert, T., Panadero, E., & Schellens, T. (2018). Anonymity as an Instructional Scaffold in Peer Assessment: Its Effects on Peer Feedback Quality and Evolution in Students' Perceptions about Peer Assessment Skills. *European Journal of Psychology of Education, 33,* 75-99. https://doi.org/10.1007/s10212-017-0339-8

Ruan, Y., & Zhang, A. (2014). Basic Principles of Evaluation for Physics Inquiry-Based Teaching. *Teaching and Management, No. 19,* 67-68.

Sadler, D. R. (1989). Formative Assessment and the Design of Instructional Systems. *Instructional Science, 18,* 119-144. https://doi.org/10.1007/BF00117714

Saunders-Stewart, K. S., Gyles, P. D., & Shore, B. M. (2012). Student Outcomes in Inquiry

Instruction: A Literature-Derived Inventory. *Journal of Advanced Academics, 23,* 5-31. https://doi.org/10.1177/1932202X11429860

Schreiber, N., Theyßen, H., & Schecker, H. (2016). Process-Oriented and Product-Oriented Assessment of Experimental Skills in Physics: A Comparison. In *Insights from Research in Science Teaching and Learning* (pp. 29-43). Springer. https://doi.org/10.1007/978-3-319-20074-3_3

Shymansky, J. A., Hedges, L. V., & Woodworth, G. (1990). A Reassessment of the Effects of Inquiry-Based Science Curricula of the 60's on Student Performance. *Journal of Research in Science Teaching, 27,* 127-144. https://doi.org/10.1002/tea.3660270205

Sun, J., & Zheng, C. (2021). International Research on Assessment of Autonomous Learning Ability: Current Situation, Trend and Enlightenment. *Comparative Education Review, No. 1,* 67-84.
https://kns.cnki.net/kcms/detail/31.2173.g4.20210205.1626.002.html

Swanson, D. B., Case, S. M., & van der Vleuten, C. P. (1991). Strategies for Student Assessment. In D. Boud, & G. Feletti (Eds.), *The Challenge of Problem Based Learning* (pp. 260-273). Kogan Page.

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing Evaluative Judgement: Enabling Students to Make Decisions about the Quality of Work. *Higher Education, 76,* 467-481. https://doi.org/10.1007/s10734-017-0220-3

Topping, K. J. (2018). *Using Peer Assessment to Inspire Reflection and Learning.* Routledge. https://doi.org/10.4324/9781351256889

Tsai, Y.-C., & Chuang, M.-T. (2013). Fostering Revision of Argumentative Writing through Structured Peer Assessment. *Perceptual and Motor Skills, 116,* 210-221.
https://doi.org/10.2466/10.23.PMS.116.1.210-221

Wang, J., He, C., & Zhang, M. (2010). The Effectiveness and Evaluation of Inquiry-Based Teaching. *Education Theory and Practice, 30,* 47-48+54.

Wanner, T., & Palmer, E. (2018). Formative Self- and Peer Assessment for Improved Student Learning: The Crucial Factors of Design, Teacher Participation and Feedback. *Assessment & Evaluation in Higher Education, 43,* 1032-1047.
https://doi.org/10.1080/02602938.2018.1427698

Weaver, F. S. (1989). *Promoting Inquiry in Undergraduate Learning.* Jossey-Bass.

Webb, N. M. (2009). The Teacher's Role in Promoting Collaborative Dialogue in the Classroom. *British Journal of Educational Psychology, 79,* 1-28.
https://doi.org/10.1348/000709908X380772

Webb, N. M., Franke, M. L., Ing, M., Chan, A., De, T., Freund, D., & Battey, D. (2008). The Role of Teacher Instructional Practices in Student Collaboration. *Contemporary Educational Psychology, 33,* 360-381. https://doi.org/10.1016/j.cedpsych.2008.05.003

Wimer, J. W., Ridenour, C. S., Thomas, K., & Place, A. W. (2001). Higher Order Teacher Questioning of Boys and Girls in Elementary Mathematics Classrooms. *The Journal of Educational Research, 95,* 84-92. https://doi.org/10.1080/00220670109596576

Wu, L., Yu, S., Pian, Y., & Liu, Y. (2022). Research on Collaborative Problem-Solving Ability Assessment Based on Students' Process Performance. *China Distance Education, No. 7,* 87-96.

Yan, Z., & Brown, G. T. (2017). A Cyclical Self-Assessment Process: Towards a Model of How Students Engage in Self-Assessment. *Assessment & Evaluation in Higher Education, 42,* 1247-1262. https://doi.org/10.1080/02602938.2016.1260091

Yan, Z., & Carless, D. (2022). Self-Assessment Is about More than Self: The Enabling Role of Feedback Literacy. *Assessment & Evaluation in Higher Education, 47,* 1116-1128.

https://doi.org/10.1080/02602938.2021.2001431

Yang, K., He, W., & Zhang, H. (2017). Teaching Process Mechanism Map: An Important Mediator for Understanding Teaching. *Educational Research and Experiment, 38,* 15-20+27.

Zimmerman, B. J., & Pons, M. M. (1986). Development of a Structured Interview for Assessing Student Use of Self-Regulated Learning Strategies. *American Educational Research Journal, 23,* 614-628. https://doi.org/10.3102/00028312023004614