

## Self-assessment of secondary school Islamic education teacher: validity and reliability of qualitative study

Azwani Masuwai, Hafizhah Zulkifli, Mohd Isa Hamzah

Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Malaysia

### Article Info

#### Article history:

Received Nov 30, 2023

Revised Sep 4, 2024

Accepted Sep 17, 2024

#### Keywords:

Islamic education teacher

Qualitative study

Reliability

Self-assessment

Validity

### ABSTRACT

Issues concerning the validity of qualitative studies are frequently debated among scholars. Interview techniques often lack standardized procedures, which can affect reliability, and eliminating biases in interviews is challenging. Ensuring reliability and validity is crucial for establishing research credibility and high measurement reliability. This study was conducted to explore the reliability and validity of qualitative study of self-assessment concept among Islamic education teachers towards continuous professionalism development. The qualitative study was involved six participants selected based on the purposive sampling technique using a semi-structured interview. This study aimed to validate interview data through participant verification, ensure the validity of qualitative themes through expert validation, and enhance reliability via data triangulation. Multiple methods were employed to achieve these goals, including interview protocol verification, pre-field study, data triangulation, field notes, participant verification, expert validation, and agreement assessed using the Cohen Kappa Index, along with an extended study period. The findings indicated that the study met its objectives for validity and reliability according to the recommended strategies and techniques. The use of various validation methods and attention to language barriers contributed to the study's robustness and offers valuable guidance for future research.

*This is an open access article under the [CC BY-SA](#) license.*



### Corresponding Author:

Hafizhah Zulkifli

Faculty of Education, Universiti Kebangsaan Malaysia

Bangi-43600, Selangor, Malaysia

Email: hafizhah\_zulkifli@ukm.edu.my

## 1. INTRODUCTION

Self-assessment among Islamic education teachers (IET) is a catalyst towards teachers' continuous professional development. Self-assessment also known as reflection and defined as an evaluation process of a person on himself by looking at things that have happened. Schön [1] stated that reflective thinking is a necessity by examine the past to make changes today. Haertel [2] mentioned that self-assessment is a process of collecting data and information about teachers' self-achievement in teaching. Boud and Falchikov [3] also noted that self-assessment can provoke individuals to actively engage with and reflect on their own progress and understanding. A person can identify the weaknesses and deficiencies through self-assessment and also recognize the strengths that useful to help teacher to make any plans and action from it. Mezirow [4] mentioned that reflection is a persons' belief to respond on something that encourages them to think constantly, be enthusiastic and always respond. Halim *et al.* [5] stated that the study of the concept of teacher evaluation and assessment is essential because their beliefs or views on that matter will affect the practice in the classroom.

Self-assessment is an important aspect in the development of teaching [6], [7] which needs to be taken into account by a teacher that can contribute to his professionalism. Sedikides and Strube [8] stated that self-assessment is important to allow a person to realize the way or steps that need to be taken to improve themselves. Nor'ain *et al.* [9] also stated that critical reflection is the driving force for teacher change and renewal. The good of reflection process is able to develop the level of teacher professionalism. Previous study [6] proved that self-assessment has significant positive contribution to professional behavior or efforts to develop teacher professionalism, which was 35.1% ( $p < 0.05$ ). Howard [10] also stated that continuous teacher reflection is one of the requirements in teacher education towards professionalism. Haertel [2] also added that self-assessment will add data to the teacher's achievement.

However, there are several issues arise in self-assessment and teacher professionalism. Firstly, self-assessment is rarely applied by teachers. McNamara and Burton [11] stated that the study about teaching behavior and self-assessment is still not carried out even though it is important in improving education. Yüce and Mirici [12] also revealed that teachers do not implement self-assessment in the context of the real class even they had admitted that they like it to be implemented. This statement was once voiced by Iwanicki and Mceachern [13] that self-assessment is rarely used even though it is an effective technique in identifying teacher development needs. This statement supported by Rozlan [14] with the highlights of the review of self-assessment study, it can be concluded that reflective practice in education still considered low-used although it is benefits to improve teaching and knowledge acquisition. Secondly, issues related to self-assessment researches that have been carried out were give an inaccurate assessment. This was stated by Vejvoda *et al.* [15] that self-assessment is often criticized for being inaccurate and bias. This statement supported by Ernst *et al.* [16] which stated that the accuracy of teachers' self-assessment is little known especially related to individual self-assessment.

Thirdly, lack of research related to self-assessment among teachers, as Muhamad [17] stated that most studies towards the formation of self-concept was focused more on students than teachers. Fourth, there are issues related to the low professionalism of religious teachers had identified. Study by Rahman *et al.* [18] found that the level of quality of alumni AIAS-PNU of religious teachers were at very worrying level. Highest assessment for teacher quality items reached relatively low percentage for good evaluation. For example, the highest rating (good) for moral practice item is 40%, intellectual skills item (26%), social and communication items (60%), knowledge items (55%) and trust and responsibility items (43%).

Fifth, issue related to professionalism can also be seen on the low level of professionalism of teachers. Patah and Boon [19] stated that the low level of teacher professionalism caused by unbalanced variable definition of professionalism. Cramer *et al.* [20] also noted that teacher professionalism is conceptualized differently in theoretical terms. This view can be found in a study by Liu *et al.* [21] on knowledge, skills, and tendencies of a teacher in their professional practice. There are also contradictions in the perception of teachers and society towards teacher professionalism which is significant [22]. The contradiction also found in management of teacher professionalism which some of them are exam-oriented or quality oriented [23]. The implementation of teacher lesson plans that caused from the notion of professionalism, accounting for teacher agency, self-efficacy and autonomy of the teacher himself also existed the inconsistency [24]. Finally, the issue related to the knowledge and perception of the teacher professionalism is low. Zhang and Liu [23] identified three patterns teacher's reaction to professionalism namely 'Why bother?', 'Fight', and 'No meaning' which overall showed a lack of perception of encouraging professionalism. Keshmiri *et al.* [25] also found that adherence to professional values among educators is low which is less than 50% ( $p = 0.002$ ) compared to accountability towards professional tasks. This situation also proved by Hanifah *et al.* [26] that found that the correlation analysis between teachers' understanding and knowledge and values of teacher professionalism were moderate which are knowledge ( $r = 0.392$ ) and value ( $r = 0.300$ ). It can be concluded, from the issues and problems stated showed that self-assessment and continuous professional development of teachers are two important things to ensure an educational growth.

Consequently, teachers need to equip themselves with reflective strategies as a priority for professional development [11] which is a necessity for improvement and meaningful self-change [27]. Teachers need to aware to improve their skills and competency in order to achieve the expected level of professionalism. As emphasized by Mutmainah [28], self-assessment is an effort that able to drive people towards loving Allah SWT more, become someone better and able to achieve more meaningful life. Self-assessment was considered successful if it is related to the teacher's own needs. Therefore, the researchers tend to explore the concept of self-assessment in the context of IET by studying at its definition, purpose, the aspects that need to be observed, the appropriate time to carry out self-assessment and what are the possible obstacles that arise to practice it. The IETs point of view about self-assessment is important to give the real picture of its implementation as it is a phenomenon study and for the literature itself.

## 2. METHOD

### 2.1. Research design

A qualitative research study is a phenomenon study that focuses on a process and event actively in building the reality and meaning of culture holistically. Based on qualitative methods, this type of study is process-oriented and more concerned with the dynamic reality of the results of the study. The qualitative analysis carried out is based on thematic analysis and uses inductive logic, which is a process of thinking from general things to specific things.

Observations in the context of qualitative study are natural and independent and focus on only a few cases and subjects based on a holistic paradigm. Researchers of qualitative study are also closer to the data where any perspective involves insiders or the researcher himself is directly involved in the study [29]. Qualitative researchers need to test and demonstrate that their research is reliable. Credibility in quantitative research depends on the construction of the instrument, but in qualitative research the researcher himself is the instrument [30]. According to Piaw [31], there are three conditions that determine the level of validity and reliability in the measurement of research studies, namely: i) untrustworthy and invalid is low measurement reliability and low measurement validity; ii) believable but invalid is high measurement reliability and low measurement validity; and iii) believable and valid is high measurement reliability and high measurement validity. Among these three situations, the third option needs to be reached by the researcher who explains that the measurement on the same subject, using the same measuring device, the collected values are almost identical, all correspond to the true value of the measured variable; measurement results are very meaningful; and the factors of research and accurate measurement methods used.

Scholars frequently discuss issues related to the validity of qualitative studies. While evaluating a single set of qualitative studies is permissible [32], relying on just one technique can lead to errors associated with that method. For instance, interview techniques often lack standardized procedures, which can impact reliability [33]. Furthermore, Robson [33] noted that it is challenging to eliminate biases or influences in interviews. To address these issues, Patton [30] recommended using a combination of open and semi-structured interviews. Additionally, to keep interviews focused on the study's context, researchers should prepare a list of general questions related to the topic under investigation [30].

Given the focus of this study on self-assessment, the researcher selected a qualitative approach as the most suitable method to thoroughly investigate the phenomenon among IETs. Participants were chosen using purposive sampling techniques, targeting individuals with extensive experience in the field who met specific criteria and possessed relevant information related to the phenomenon under study [34], [35]. This research will concentrate on a qualitative study, with a focus on validity and reliability [31], and will also address the ethical considerations crucial for strengthening the study's credibility. Specifically, the objectives of this study are: i) to obtain the validity of the interview data from the verification of the participants; ii) to obtain the validity of the theme of qualitative data from expert validation; and iii) to obtain the reliability of interview data with data triangulation.

The intended audience for this study includes researchers using a qualitative approach and those employing interview techniques for data collection. This audience can benefit from understanding the necessary steps and strategies to achieve validity and reliability in their data. Additionally, focusing on the credibility of research data will enhance their comprehension of validity and reliability aspects.

### 2.2. Validity and reliability

Trochim [36] described reliability as the study's ability to achieve consistent results when the same measurement is administered repeatedly. However, reliability is difficult to determine accurately in qualitative research because human behavior is not static even in the same situation when replicating or repeating the study. Therefore, it can be understood that in the context of qualitative research, the concept of reliability is related to the issue of reliability and consistency of data collected where the main question is not that the same findings will be found again but that the results or findings of the study are consistent with the data collected [37]. This is closely related to the concept of quality research where reliability is a concept that assesses quality in quantitative studies with the "purpose of explaining" while the concept of quality in qualitative studies aims to "generate understanding" [38]. Stenbacka [38] view in discussing the concept of reliability in qualitative research is in line with the view of other researchers [31], [37], [39] who assert that the difference in the purpose of assessing the quality of research in quantitative and qualitative research is one of the reasons that makes the concept of reliability irrelevant in research qualitative.

However, there are also several researchers who support the concept of reliability in qualitative studies [30], [40], [41] who insist that validity and reliability are two concepts that are interrelated and important when designing a study, analyzing data and in analyzing the quality of a study. To make it more specific, the term reliability is entirely reserved for quantitative studies. Whereas, for qualitative studies, the term dependency is used to refer to the concept of reliability [41].

In qualitative research, some scholars argue that the term “validity” may not apply, but they recognize the need to develop suitable methods for assessing qualitative studies. Creswell and Miller [42] suggest that a researcher’s perception of validity and their chosen paradigm influence how validity is understood. Consequently, researchers have developed their own interpretations of validity and introduced various terms to describe it. Commonly used concepts include quality, rigor, trustworthiness, credibility, dependability, and confirmability [34], [38], [41], [43], [44].

Validity in the qualitative paradigm has been defined as the extent to which research findings accurately describe the real situation and are supported by evidence, ensuring they are true and certain [34]. Qualitative research can be considered trustworthy if others, even those who only read the research report, recognize the findings as reflective of their own experiences [31]. This concept is referred to as external validity, which pertains to the generalizability of research findings to other situations [37]. To enhance the reliability of qualitative research, Piaw [31] suggests several key strategies for researchers to consider. These include employing triangulation methods, utilizing trained researchers, implementing peer review, and using note-taking tools as evidence. Detailed descriptions of these strategies are provided in Table 1.

Reliability and validity in research are essential for ensuring that the data is robust, replicable, and the results are accurate, thus maintaining the integrity and quality of the measurement instrument [45]. Creswell [46] highlighted that validity and reliability are key factors in determining an instrument’s suitability and usability. Piaw [31] defines validity as the measurement’s ability to accurately capture the true value of a concept within a hypothesis. This view aligns with Kerlinger and Lee [47], who asserts that validity measures the accuracy of a measurement tool in reflecting the variable it is intended to assess.

Table 1. Techniques to increase the reliability of qualitative studies

No	Techniques	Explanation	Example
1	Using the triangulation method	Using several different researchers Do research at different times Do research in different places	In an interview session, several interviewers are used to ask the same questions at different intervals
2	Using a trained researcher	Involve trained researchers to make careful and systematic observations and reports	
3	Using peer review	Comparing the research data recorded by the researcher and the opinions of peers about the subject	
4	Using a recorder as evidence	A recording tool as proof of the reliability of the study that records the results of the interview/observation session	Interview forms, cassette recorders, video tapes and so on

### 2.2.1. Validity

Validity, according to Campbell and Fiske [48], is the agreement between two attempts to measure the same characteristic maximally with different methods. Validity also according to some other writers can be defined with several specific meanings. Among them is related to research instruments from which data can be inferred, which are instruments that have: i) appropriateness [49], [50]; ii) truthfulness [51]–[53]; iii) meaningfulness [54], [55]; and iv) usefulness [56], [57].

Validity measures the accuracy of an instrument in a study, ensuring that it captures all relevant features or concepts being assessed [30]. For an instrument, the purpose of ensuring validity is to guarantee that the items included accurately reflect the concept being measured: i) defensibility, because the results of the study are accurate and useful [58], [59]; ii) accuracy, that is in answering the research question [60], [61]; iii) appropriateness, which is relevant to the purpose of the study; iv) meaningfulness, that is giving meaning to data through scores [54], [55]; and v) usefulness, that is being able to make decisions in relation to what is being sought or produced because the results of the assessment can provide meaningful information about the topic and variables to be measured to infer research data [56], [57].

In this study, the researcher has considered both internal and external validity. Internal validity is a level at which the conclusions/inferences made by the researcher about the relationship between variables and events can be trusted. McMillan *et al.* [62] stated that it refers more to the general concept, adjustment between the categories and interpretations made by the researcher with what actually happened. The increase in the determination of internal validity used by the researcher for this study is the triangulation method (multipurpose methods of data collection). It is used when collecting data from multiple sources, such as document analysis and individual interviews.

The method of triangulation used with the aim of researchers can research a phenomenon from various perspectives. It not only adds evidence, validates data and confidence in the consistency of information obtained, but also supports data from various sources. The used of this method can help researchers to research and obtain various information from various perspectives [30], [34], [63]. For example, the researcher triangulated the interview data by comparing responses from different respondents

and informants, and by integrating findings with document analysis. Additionally, triangulation was performed across multiple research sites and combined qualitative data with quantitative data.

Transferability is an empirical matter that depends on the degree of similarity between the context of delivery and reception [41]. Transferability inferences cannot be made by researchers because they only know the context of the presentation. What the researcher can do is to generalize from the sample to the population, or from one setting to another. Researchers can also transfer inferences/summaries from contexts (certain settings) to other settings [64].

In this study, the researcher employed an inductive approach, collecting data from specific examples to develop a theory or abstract concept. This generalization process extends from particular settings to broader contexts, helping to define constructs more clearly [64]. For instance, the researcher identified key aspects related to self-assessment to construct the study's theoretical framework. The interviews revealed that these abstract concepts could be further expanded and refined for more precise definitions. To enhance external validity, the researchers utilized several strategies recommended by Merriam [34] and Lincoln [41], including comparability, translatability, and member checking.

The comparison involves the characteristics of the participants, the documents used, the analysis and concepts used in the study were become an information for researcher for further studies. Therefore, the researcher will ask the study participants in the interview to define each construct presented in order to get a clearer meaning. The translation strategy is to explain the theoretical framework used in the study clearly and in detail so that other researchers can use the same framework in further studies in the future.

A peer review strategy was used by reconnecting with the interview study participants to check the validity of the data interpretation through the previous interview process and obtaining their consent [44]. This strategy is aimed at verifying information to increase accuracy, credibility, validity and transferability of information. The step taken by the researcher is to contact interview respondents to check and ensure the accuracy of all the information and facts provided by them through the member check form. It is carried out with the aim of respondents can check the data and preliminary interpretations to ensure the accuracy and appropriateness of the findings throughout the study. Peer check evaluation is used where fellow researchers in the field of Islamic education check the interview data to ensure that the interpretation made by the researcher is in line with other individuals.

Member checking can be done either throughout the investigation or at the conclusion stage to achieve a credible review [64]. At this stage, the researcher will ask friends who are in the scene to make an analytical review related to categories, conclusions and the researcher's interpretation of the subject being studied. Spradley [65] suggested that structured questions be asked to informants to confirm the analytical domain that has been constructed by the researcher. If the informant expresses agreement with the interpretation of the following researcher, it is proof of the credibility of a decision or finding obtained.

### 2.2.2. Reliability

Reliability was defined as a measurement value that can be used as a guide to determine the consistency of an item's score [66]. It is an important element in the concept of reliability which means that if the same item is tested several times to the same subject, then the result/answer score given remains the same or almost the same. It also refers to the stability of a measure or measuring tool or study or questionnaire across time against an idea [34]. One of the main purposes of achieving reliability is to find out whether the measure gives the same answer when it is used to measure the same concept to the same population or sample or respondent [47], [67]. In qualitative research, reliability is the extent to which the researcher can be trusted by others due to integrity, honesty (trustworthiness) and personality shown [26]. It is also said to be dependability which is the extent to which a study has integrity and honesty shown to encourage other researchers to depend on it and the study can be repeated.

The reliability of the qualitative data in this study is related to the researcher's observations about the IET whether it has internal or external consistency. Internal consistency refers to the data obtained organized in a meaningful way. While external consistency is confirmed by checking observations obtained with various other data sources [68]. The reliability used by the researcher in this qualitative study involves several steps as suggested by Creswell and Poth [35], namely verification of semi-structured question inventory by supervisors and experts in the field, pilot studies, data triangulation, field note reports and diary, expert verification of themes which is built, a long period of time in the study. Finally, the Cohen Kappa calculation of the level of expert agreement on the theme was also performed.

### 2.3. Sample

The sample for the pre-field study was conducted with IET. While for the actual study, the study participants have been determined using purposive sampling techniques with the participant selection criteria that have been set. The advantage of this technique is practical because the individuals being studied are suitable and qualified and not much time is needed to gain the trust of the study participants [69]. Thus, the

use of this technique can increase the high degree of reliability whereas the participants were selected based on the respondent's own ability, willingness, and readiness to provide any information needed to answer research questions. Based on these established criteria, six study participants were selected. With six interviews conducted, the conclusion has been drawn and the data collected is saturated.

## 2.4. Instrumentation

### 2.4.1. Preparation of interview protocol

Most qualitative studies in the field of education have make pre-field studies as one of the important aspects of the study [70]–[73]. The purpose of pre-field study in qualitative research is to identify the appropriateness of the interview instrument used for the respondents. In this study, a semi-structured interview instrument was prepared before being handed over to the supervisor to see the alignment between the research objectives and themes with the research questions. This step of validity is very important as one of the validations in qualitative research [74]. The verification of supervisors and colleagues in this study is also one of the forms of data reliability. One of validity of qualitative data is through verification of supervisors and fellow researchers on the regularity of the research conducted [67]. The preparation of interview protocol questions was done several times until it really met the requirements of the study. Draft of interview protocol question construction was done since May 11th, 2022 until it was confirmed by the supervisor on June 3rd, 2022 to be taken to the expert evaluation stage.

### 2.4.2. Validation of interview protocol

At the expert validation stage, two panels of experts, one comprising Islamic education specialists and the other consisting of qualitative research experts were appointed. An official email was sent to these experts, which included several attached documents for the validation process. These attachments included the interview protocol expert review validity form and the interview protocol validity review declaration form. After receiving feedback from both expert panels, the comments and suggestions were reviewed with the supervisor to refine the interview protocol questions. By June 17th, 2022, the interview protocol had been fully revised and was ready for the pilot study. Table 2 details the chronology of the preparation and verification of the interview protocol questions. Table 3 displays the interview questions used for expert validation, while Table 4 lists the final research interview questions.

## 2.5. Data collection

### 2.5.1. Ethics of conducting interviews

Typically, study participants and the researcher must agree on the study's objectives to ensure alignment between the aspects analyzed and the participants' perspectives [75]. This consent is crucial for making participants aware of the study's goals and the data needed, as well as for providing them with preliminary guidance, ensuring their comfort, and achieving high-quality data. For this study, the researcher prepared a research participant consent form, which was provided to participants before the interview sessions began. This form also serves as a contract between the researcher and the participants for data collection purposes.

The study participant consent form includes several key details: the researcher's self-identification, the background of the study, a description of the study procedures, and a statement on confidentiality ethics. It also provides space for participants to ask questions and includes contact information for the researcher. Additionally, the form features an acknowledgment section where both the participant and researcher confirm their agreement.

Table 2. Preparation and validation of the interview protocol chronology

Date	Matter	Notes
11th May 2022	Draft first interview protocol questions	Towards the construction of the study
12th May 2022	Draft second interview protocol questions	Directed to the research question
26th May 2022	Draft third interview protocol questions	Aiming at the phenomenon of study
3rd June 2022	Draft fourth interview protocol questions	Confirmed by supervisor
6th June 2022	Expert evaluation for the validity of the interview protocol	Expert 1: Islamic education expert Expert 2: Qualitative research expert
10th June 2022	Acceptance of expert confirmation 1	Correction of the sentence structure in the question.
13th June 2022	Acceptance of expert confirmation 2	State the meaning of the construct so that the study participants have clear understandings.
17th June 2022	Expert evaluation discussion	Rearrangement of questions and sentence structure in questions according to expert comments.

Table 3. Interview questions for expert verification

No.	Questions
1	What is meant by self-assessment?
2	What are the aspects of self-assessment?
3	What is the scope of self-assessment in the teaching aspect?
4	What are the constructs that need to be in self-assessment?
5	What items should be included in the self-assessment?
6	What is the appropriate form of question used to measure self-assessment?

Table 4. Interview questions for the study

Part of questions	Questions
Opening questions	<ul style="list-style-type: none"> <li>– Assalamualaikum (Greetings)</li> <li>– Thank you Prof./Dato'/Dr./Madam/Sir for the time to share your experience and expertise.</li> <li>– I am in the process of developing a self-assessment instrument among Islamic Education Teachers in Secondary School. This interview is part of the process in obtaining information for that purpose.</li> </ul>
Introduction questions	<ul style="list-style-type: none"> <li>– How long has Prof./Dato'/Dr./Madam/Sir been teaching/working in this field?</li> <li>– Prof./Dato'/Dr./Madam/Sir was selected for the interview because Prof./Dato'/Dr./Madam/Sir has been identified as an expert in the field of Islamic Education.</li> <li>– Can you tell me the experience of Prof./Dato'/Dr./Madam/Sir in this field?</li> <li>– Today's interview focuses on self-assessment among Islamic Education teachers secondary schools. Has Prof./Dato'/Dr./Madam/Sir ever practiced self-assessment?</li> <li>– What is the meaning of self-assessment according to Prof./Dato'/Dr./Madam/Sir?</li> </ul>
Transitional questions	<ul style="list-style-type: none"> <li>– Teaching at school or in the education field, does the Prof./Dato'/Dr./Madam/Sir take into account the aspect of self-assessment? Can you describe?</li> <li>– When is the self-assessment usually carried out? Can you explain?</li> <li>– How was it carried out? Can you describe?</li> <li>– Where was it carried out? Can you describe?</li> </ul>
Key questions	<ul style="list-style-type: none"> <li>– What are the aspects that need to be observed for teacher in practicing self-assessment?</li> <li>– What is the scope/element of self-assessment in the teaching profession that teacher can do?</li> <li>– What forms of behavior can often be observed from the teacher who practicing self-assessment?</li> <li>– What are the types of constructs/items that Prof./Dato'/Dr./Madam/Sir used/that can be used in self-assessment?</li> <li>– How do you give a score/level in the self-assessment that has been done?</li> </ul>
Closing questions	<ul style="list-style-type: none"> <li>– What are the problems encountered/may occur when practicing self-assessment?</li> <li>– The interviewer concludes the purpose of the interview and then asks the question: Is there any information/statement that Prof./Dato'/Dr./Madam/Sir would like to add?</li> </ul>

#### a. Pre-field study

On July 5th, 2022, the researcher conducted a pre-field study with an Islamic education expert via Google Meet. The session was recorded using the Google Meet recording feature and a mobile phone, and the researcher took notes for reference. The pre-field study results, reviewed by the researcher and approved by the supervisor, indicated alignment between the responses and the study's objectives. However, the researcher identified the need to enhance the questioning technique by considering the respondent's descriptions and improving probing strategies.

The results of the pre-field study, along with the field note report, were submitted to the supervisor for review to ensure that the interview results aligned with the study's objectives. The supervisor also assessed how the researcher documented the field notes. Following this review, the researcher received approval from the supervisor to proceed with the actual study.

#### b. Field study (actual study)

After the pre-field study has been conducted, it is followed up with an actual study in the field starts with the determination of study participants which uses purposive sampling techniques. The criteria for the selection of participants were set as stated by Merriam [34]. The four criteria set for the selection of study participants are as: i) experienced in the field of Islamic education; ii) experienced in the field for over 15 years; iii) teaching subjects related to Islamic education; and iv) approachable, friendly, and cooperative. Based on these established criteria, six study participants were selected and willing to cooperate for this actual study process. A summary of the demographics of the study participants can be seen in Table 5.

Before conducting the interview sessions, the researcher obtained preliminary consent from the study participants. An official email was sent to each participant detailing the intention to conduct the interview and including several attachments: the appointment letter, a study synopsis, the interview protocol instrument, and the respondent consent form. After receiving consent, the researcher contacted the participants via telephone, message, or email to schedule a mutually agreed-upon date, time, and location for the interviews. Once all details were confirmed, the interviews were conducted in a well-organized manner.

Table 5. Study participant demographics

Participant	Nickname	Experience in Islamic education	Experience period
Participant 1	Teacher Rabiah	Excellent Teacher of Islamic Education; Chairman of the Committee; Preaching Trainer; SPM Exam Paper Markers; Area Appraiser; Item Drafter	26 years
Participant 2	Teacher Norzihan	Tutor; Senior Lecturer; Deputy Dean	17 years
Participant 3	Teacher Maria	Senior Assistant Curriculum; Evening Senior Assistant; Senior Teacher of Humanities; Chairman of the Committee; Committee of the Islamic Princess Movement; Islamic Education Trainer; Co-curricular Technical Officer	31 years
Participant 4	Teacher Noriah	Trainer; SPM Exam Paper Markers; Area Appraiser; Item Drafter	
		Co-Curriculum Trainer; Secretary of Referee; Secretary of the Islamic Education Committee; ADNI Deputy Director	27 years
Participant 5	Teacher Fauziah	Tutor; Senior Lecturer; Head of Study Center	20 years
Participant 6	Teacher Ros	Media Teacher; Special Education Teacher; Senior Lecturer	25 years

## 2.6. Data analysis

The interview data was analyzed using thematic analysis [34]. The validation of themes by experts was assessed using Cohen Kappa Index [76]. This study specifically concentrates on the theme verification process conducted by appointed experts, focusing on the validity and reliability of qualitative research.

## 3. RESULTS AND DISCUSSION

### 3.1. To obtain the validity of the interview data from the verification of the study participants

One of the ways to increase the validity and reliability of the primary qualitative data collected by the researcher is through verification of the collected interview data [67], [77], [78]. The recording conducted by the researcher involves two forms of recording, namely voice recording and video recording. However, this video recording was only recorded after the researcher got permission from the participants. Video recording can be made using Google Meet platform that arranged between the researcher and the participants. While face-to-face meetings, there is no video recording was made. Nevertheless, voice recordings were made for all interview sessions conducted.

The purpose of the video recording is to see facial expressions in response to the questions. This video recording is to support the voice recording so that the researcher cares about each participant's answer. The voice recording for each participant was then transcribed right after the interview ended. Therefore, all data that support the recording could be reported in the field notes. The interview session conducted from July 2022 until September 2022 which took three months altogether. The interviews conducted have achieved sufficient information efficiently according to proper ethical considerations of the period.

Once the transcript is complete, the researcher provides a copy to the participants for review. Participants are given the opportunity to re-read the conversation and correct any errors, as well as to sign the transcript for verification. The researcher includes a space for participants to evaluate and sign the document. This process of fact-checking, acknowledgment, and confirmation by participants enhances the validity and reliability of the interview data. Table 6 displays the schedule of the interview sessions and the dates when transcripts were prepared for each session.

The procedures followed included various techniques such as voice and video recording, transcribing immediately after interviews, and documenting data in field notes. These steps were carried out with careful attention to ethical considerations and participant verification. The strategy aimed to enhance the accuracy, credibility, validity, and transferability of the information, ultimately ensuring the validity of the interview data.

Table 6. Interview session and transcript preparation

interview Date	Participant	Platform	Transcript preparation date
5th July 2022	Participant 1	Google Meet	24th July 2022
19th August 2022	Participant 3	SMK Subang, Shah Alam	23rd August 2022
22nd August 2022	Participant 2	Google Meet	26th August 2022
25th August 2022	Participant 4	SK Methodist ACS Pelabuhan Klang, Klang	27th August 2022
30th August 2022	Participant 5	Google Meet	30th August 2022
8th September 2022	Participant 6	Teacher Education Institute, Bangi Islamic Education Campus	12th September 2022

### 3.2. To obtain the validity of the theme of qualitative data from expert validation

After confirming and organizing all transcripts, field notes, and diary reports, the researcher began managing the data to identify themes from the In-Vivo study using inductive approach. Table 7 displays the six themes identified from the analysis. Once the data was processed and patterns in the study findings were



established, the researcher prepared a set of expert consent forms for the constructed themes. Expert agreement on these themes is crucial [79]. In Table 7, the Cohen Kappa Level of agreement indicator shown by the experts as an inter-rater (agreement between raters). The experts are outsider who confirms the theme construction done by the researchers [79]. Analysis of the Cohen Kappa Index was used to determine the degree of agreement of the analysis unit with the theme under study [80]. For this purpose, the researcher has obtained the evaluation of three experts in qualitative analysis in the field of Islamic education who were appointed to be inter-raters for the themes that were constructed as shown in Table 8.

The results of the expert evaluation indicated that the thematic construction by the researchers achieved a Cohen Kappa measurement level above 0.70, indicating a high level of agreement, as shown in Table 9. This high Cohen Kappa score provides strong evidence of the research data's reliability. Additionally, the analysis units used by the researcher align well with the themes, as confirmed by the expert agreement. The expert validation process is vital for ensuring the reliability of the study. All themes identified during the analysis were reviewed and confirmed by experts, following the recommendations of Creswell and Poth [68], and achieved a high level of agreement. As a result, the validity of the qualitative data themes, as validated by experts, was established.

Table 7. Themes of the study

No.	Theme
1	Meaning of self-assessment
2	Purpose of self-assessment
3	Aspects of self-assessment
4	Time for conducting self-assessment
5	Qualities of teacher who practice self-assessment
6	Challenges encountered in self-assessment

Table 8. List of expert panels of agreement analysis of qualitative data theme construction

No.	Expert	Representatives	Position and qualifications	Expertise	Level of agreement (Cohen Kappa)
1	Expert A	University	Academic Deputy Rector	Islamic pedagogy and education	0.86
3	Expert B	School	Excellent Teacher	Assessment and measurement, Islamic education	0.84
2	Expert C	Educator's institution	Lecturer	Al-Quran and as-Sunnah education and Islamic education	0.81

Table 9. Cohen Kappa level of agreement indicator

Indicator	Value
Perfect agreement	0.81-1.00
Substantial	0.61-0.80
Moderate	0.41-0.60
Slight	0.00-0.40

### 3.3. To obtain the reliability of interview data with data triangulation

One of the ways to make high confidence of qualitative data is to implement the triangulation between the data [81], [82]. Triangulation between the data in this study was chosen by several methods to reaches high level of confidence of the data collected. The methods chosen are: i) the triangulation of multi-field studies (multiple case or multisite studies); ii) the triangulation between interview data with IET; iii) the triangulation between interview data of participants with experts (consist of administrators and academic experts); iv) the triangulation between interview data with observation and document analysis; and v) the triangulation between qualitative and quantitative data.

According to Scott [83], the reliability of data collected from interviews is made through systematic cross-referencing of information provided by informants on the same line. Reliability was achieved when the same information is repeated several times in one line from one informant and the researcher can use this technique during the data coding process through interview transcription. Triangulation between various fields (multiple case or multisite studies) is one form of data triangulation with high reliability [34]. As the study of various cases was a study that involved various places of the field study. Data collected and analyzed were come from various field sites. According to Merriam [34], this kind of data can further increase the external validity or generalization of research findings.

#### 3.3.1. Data triangulation between participant and informant interviews

Triangulation between interview data of the main participants, i.e. IET with support participants' interviews or informants. The triangulation consists of: i) IET key data with the administration; ii) IET key

*Self-assessment of secondary school Islamic education teacher: validity and reliability ... (Azwani Masuwai)*

data with policy makers; and iii) IET key data with academic experts. During the study, the researchers initially informed participants that their interviews would be cross-referenced with those of other informants. This approach was intended to prevent participants from exaggerating or embellishing their responses. As a result, each theme developed showed clear and high coherence, with consistency observed across the interviews, as discussed in the research findings.

### 3.3.2. Triangulation of interview data with document analysis

In addition to triangulating the interview data, the researcher also used document analysis to enhance the reliability of the data. Document analysis served as supplementary data that supported the interview findings but could not stand alone without them. This analysis included reports detailing aspects of IETs' self-assessment, referencing existing instruments, policies, Quranic verses, and previous studies. Employing triangulation with document sources helps corroborate and add evidence from multiple sources, thereby confirming and validating the information gathered from interviews and observations [84].

### 3.3.3. Field notes and diary references

The high reliability value of qualitative data also involves the effort of organizing data collection in the field [67]. Evidence of the implementation of data collection process such as appointments, formal and informal interviews, observation, and document collection are also a form of high reliability. Eventually, every implementation of data collection work in the field needs to be informed in a report that called 'field note' [85], [86]. These field notes are reinforced with a brief diary entry of the researcher.

To facilitate the creation of field notes, the researcher prepared a small book to record detailed notes for each interview session. These booklets were systematically organized into a complete report, categorized by the platforms used [86]. In addition, diaries were employed to plan the field research and serve as part of the validation process [87] to recorded all appointments, locations, and interview platforms.

The researchers employed triangulation by comparing responses from various respondents and informants and integrating these findings with document analysis. Triangulation was also applied across multiple research sites and involved combining qualitative and quantitative data. These methods aimed to gather information from different perspectives and minimize potential biases or influences in the interviews, thereby ensuring high data reliability.

## 3.4. Discussion

Based on the findings, all objectives have been met by following the recommended strategies and techniques for ensuring validity and reliability in qualitative research. To ensure the validity of the interview data through participant verification, interviews via Google Meet were video-recorded, while face-to-face interviews were audio-recorded. All recordings will be transcribed, and the transcriptions will be reviewed and approved by the participants. To ensure the validity of the themes through expert verification, the results of the theme analysis were reviewed by three experts in qualitative research. This validation process demonstrated a high level of agreement, as indicated by the Cohen Kappa Index. Meanwhile, to ensure the reliability of the interview data, data triangulation was employed to enhance confidence in the findings. This involved cross-referencing data from different participants and informants as well as comparing it with document analysis. The processes implemented were designed to guarantee that the data is true, accurate, and meaningful, making it suitable for the current research and valuable for guiding future studies.

The validity and reliability process carried out in this study is the same as most other qualitative studies involving interviews. Among them are studies conducted by Henry and Thorsen [88], particularly, the use of field notes was used to explore disaffection and lack of participation in response to language activities in an English as a foreign language (EFL) classroom. The field notes and observations of lessons in a multiple case study conducted in language classrooms in Sweden have been examined to gathered the data. Meanwhile, Batenburg *et al.* [89] had used expert validation and data triangulation between participants to identify the interaction opportunities in each chapter of interactional opportunities provided in Dutch course books for EFL. The coding scheme was checked and rechecked against samples of the data and was then used by trained experts. In study by Mohamed and Jasmi [90], the study explore teaching and learning practice among preschool Islamic education teachers, the validity of the data was obtained through triangulation of different sources of data, long duration of data collection which was six months and validated by the participants of the research. Whilst, the reliability of the data was obtained through audit trail and experts validation based on Cohen Kappa correlation validation.

Moreover, Spenader *et al.* [91] have used triangulation of interview data with document analysis to explored the use of content-based instruction (CBI) in World language teaching in the United States. The analysis was thorough, involving a number of stages, including preliminary examination of selected units, developing and verifying guidelines for the analysis, fine-tuning categories, and coding lessons based on the

categories that build from the CBI. Subsequently, in Rozali and Jasmi [92], the study to examine practical teaching and learning characteristics implemented in secondary schools following the mainstream of education, interview data of primary study participants were triangulated with interview data and secondary data from supporting study participant interviews, observations, and document analysis data.

Several methods were applied to improve the reliability of the qualitative study [31]. Specifically, triangulation was used for objective 3 to ensure the reliability of interview data. A trained researcher was involved across all objectives to enhance the validity of interview data through participant verification, expert validation of qualitative themes, and data triangulation. Additionally, peer review and recording devices were utilized for objective 1 to support the validity of interview data through participant verification.

The study emphasizes the importance of validity and reliability throughout the entire qualitative research process, from the initial data collection through to the final acquisition of data. Ensuring validity and reliability is crucial for assessing whether the data can effectively address the research questions and be useful for any future studies. Additionally, these aspects offer valuable guidance for future researchers conducting qualitative studies, particularly those involving interviews. The use of multiple validation techniques in qualitative research is important for ensuring the findings are accurate, credible, and reliable by reducing biases, verifying results from different angles, and providing a well-rounded view of the data.

For this study, all qualitative data that obtained through the validity and reliability process will be used as the basis for the construction of quantitative instruments for the next quantitative data collection process. The exploration carried out through qualitative approach earlier, will be the initial phase of the study that led to the second phase of the study which focuses on the quantitative approach. This sequential exploratory design or better known as exploratory sequential design was used for studies that followed each other over time. It prioritizes the collection and analysis of qualitative data in the first phase and the findings from it directed to the second phase which is a quantitative study that involves larger sample. The researchers will interpret how these quantitative findings are built from qualitative findings [93].

#### 4. CONCLUSION

The present study investigated the important of validity and reliability of qualitative research. Validity and reliability are crucial for ensuring trustworthiness, credibility, dependability, and confirmability in assessing both the external and internal validity of a study. They involve validating the research conclusions and inferences based on the available data. In qualitative research, various types of data can be used, and the choice of data validation methods depends on the study type and research instruments employed. The study utilized a qualitative study approach, which enabled an in-depth understanding of the self-assessment of IET for continuous professional development. The thematic analysis of qualitative data revealed six themes of self-assessment practice which are the meaning of self-assessment, the purpose of self-assessment, aspects of self-assessment, the time to carry out self-assessment, the character of teachers who practices self-assessment, and problems occur in self-assessment. For this study, multiple methods were employed to ensure data validity and reliability, although not all available options were used. However, the choice made is the best one that fulfil all the techniques suggested by expert to establish the reliability of qualitative data.

The selected methods included interview protocol verification, pre-field study, data triangulation, field notes, participant verification of interview data, expert verification, and agreement among experts measured by the Cohen Kappa Index, as well as an extended study period. Additionally, the researcher utilized several strategies to enhance validity and reliability, including conducting interviews in Malay language (*Bahasa Melayu*) to prevent issues related to meaning differences in translation. The use of multiple validation techniques and careful consideration of language barriers contributes to the robustness of the study and provides valuable guidance for future research.

#### ACKNOWLEDGEMENTS

This study was funded by National University of Malaysia, Grant Number: GP-K019312.

#### REFERENCES

- [1] D. A. Schön, *The reflective practitioner*. Routledge, 2017, doi: 10.4324/9781315237473.
- [2] G. Haertel, *A primer on teacher self-evaluation*. Washington: Publication Series, 1993.
- [3] D. Boud and N. Falchikov, *Rethinking assessment in higher education: learning for the longer term*. Routledge, 2007.
- [4] J. Mezirow, "How critical reflection triggers transformative learning," *Fostering Critical Reflection in Adulthood*, vol. 1, no. 20, pp. 1–6, 1990.
- [5] H. Abd Halim, M. I. Hamzah, and H. Zulkifli, "Secondary school teachers' conceptions of assessment: A systematic review," *International Journal of Educational Research*, vol. 124, p. 102311, 2024, doi: 10.1016/j.ijer.2023.102311.




- [6] M. Akram and S. J. Zepeda, "Development and validation of a teacher self-assessment instrument," *Journal of Research and Reflections*, vol. 9, no. 2, pp. 134–148, 2015.
- [7] F. Korthagen and A. Vasalos, "Levels in reflection: Core reflection as a means to enhance professional growth," *Teachers and Teaching: Theory and Practice*, vol. 11, no. 1, pp. 47–71, Feb. 2005, doi: 10.1080/1354060042000337093.
- [8] C. Sedikides and M. J. Strube, "Self-evaluation: to thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better," in *Advances in Experimental Social Psychology*, Elsevier, 1997, pp. 209–269, doi: 10.1016/S0065-2601(08)60018-0.
- [9] M. T. Nor'ain, A. R. Nuruhuda, M. Azwani, and I. Noraini, "Comparison between teaching practices based on teacher educators' perception and learning experiences based on student teachers' perception at higher education institution," *Research Journal in Organizational Psychology & Educational Studies*, vol. 3, no. 6, pp. 437–445, 2014.
- [10] T. C. Howard, "Culturally relevant pedagogy: ingredients for critical teacher reflection," *Theory into Practice*, vol. 42, no. 3, pp. 195–202, 2003, doi: 10.1353/tip.2003.0031.
- [11] J. McNamara and K. Burton, "Assessment of online discussion forums for law students," *Journal of University Teaching and Learning Practice*, vol. 6, no. 2, pp. 69–83, Aug. 2009, doi: 10.53761/1.6.2.6.
- [12] E. Yüce and I. H. Mirici, "Self-assessment in EFL classes of secondary education in Türkiye: the common European framework of reference for languages (CEFR)-based implementations," *Pegem Eğitim ve Öğretim Dergisi*, vol. 13, no. 1, pp. 349–359, Jan. 2022, doi: 10.47750/pegegog.13.01.38.
- [13] E. F. Iwanicki and L. Mceachern, "Using teacher self-assessment to identify staff development needs," *Journal of Teacher Education*, vol. 35, no. 2, pp. 38–41, Mar. 1984, doi: 10.1177/002248718403500209.
- [14] M. Q. M. Rozlan, "A case study of the criticality of teaching English as second language (TESL) student teacher reflective entries," Ph.D. dissertation, Universiti Pendidikan Sultan Idris, 2022.
- [15] J. Vejvoda, M. Stadler, F. Schultz-Pernice, F. Fischer, and M. Sailer, "Getting ready for teaching with digital technologies: Scenario-based self-assessment in teacher education and professional development," *Unterrichtswissenschaft*, vol. 51, no. 4, pp. 511–532, Oct. 2023, doi: 10.1007/s42010-023-00186-x.
- [16] H. M. Ernst, J. Wittwer, and T. Voss, "Do they know what they know? Accuracy in teacher candidates' self-assessments and its influencing factors," *British Educational Research Journal*, vol. 49, no. 4, pp. 649–673, Mar. 2023, doi: 10.1002/berj.3860.
- [17] N. A. F. Muhamad, "Muslim self-concept measurement model: towards filling the research gap of self-concept among Muslims," *Islamiyyat*, vol. 45, no. 1, pp. 149–159, Jun. 2023, doi: 10.17576/islamiyyat-2023-4501-13.
- [18] A. Z. Rahman *et al.*, "Critical review of positive behavior and resilience in Islamic perspective during the COVID-19 pandemic," *Journal of Critical Reviews*, vol. 7, no. 5, 2020.
- [19] S. Abdul Patah and Y. Boon, "Issues and trends of teacher professionalism for Malaysian education," *International Journal of Academic Research in Business and Social Sciences*, vol. 11, no. 4, Apr. 2021, doi: 10.6007/ijarbss/v11-i4/9666.
- [20] C. Cramer, J. Groß Ophoff, and F. Schreiber, "Professionalität im Lehrerinnen- und Lehrerberuf," *Zeitschrift für Pädagogik*, vol. 69, no. 1, pp. 67–87, Jan. 2023, doi: 10.3262/zp2301067.
- [21] L. B. Liu, J. M. Conner, and Q. Li, "Cultivating teacher professionalism in Chinese and U.S. settings: contexts, standards, and personhood," *Education Inquiry*, vol. 14, no. 1, pp. 22–39, Jul. 2023, doi: 10.1080/20004508.2021.1950274.
- [22] J. M. S. Wong, "Are we becoming professionals? Pre-service early childhood teachers' perceptions of the professionalism of early childhood teachers in Hong Kong," *Early Years*, vol. 42, no. 1, pp. 23–38, Jul. 2022, doi: 10.1080/09575146.2021.1954604.
- [23] Y. Zhang and S. Liu, "Re-professionalisation or de-professionalisation: how do Chinese high school teachers respond to the new professionalism?" *Professional Development in Education*, vol. 50, no. 5, pp. 965–981, Jul. 2021, doi: 10.1080/19415257.2021.1959382.
- [24] C. Karousiou, M. Vrikki, and M. Evagorou, "Teachers' positioning and sense of professionalism over the implementation of cultural literacy as a dialogic and argumentative practice," *Professional Development in Education*, pp. 1–17, Jul. 2022, doi: 10.1080/19415257.2022.2095661.
- [25] F. Keshmiri, S. Jambarsang, and A. H. Mehrparvar, "Effective components of teachers' professionalism in viewpoints of various stakeholders," *Journal of Education and Health Promotion*, vol. 12, no. 1, Jan. 2023, doi: 10.4103/jehp.jehp\_1565\_21.
- [26] M. Hanifah, H. Mohmadisa, S. Yazid, N. Nasir, and N. S. Balkhis, "Professional and pedagogical competencies of geography teachers in Malaysia," *Review of International Geographical Education Online*, vol. 9, no. 2, pp. 304–318, Aug. 2019, doi: 10.33403/rigeo.509836.
- [27] T. J. Sergiovanni and R. J. Starratt, *Supervision: human perspectives*. New York: McGraw-Hill, 1983.
- [28] Mutmainah, "Muhasabah Method: Analysis of the Sufistic Psychological Approach from Al Ghazali Perspective (Concept of Spiritual Education Through Tazkiyatun nafs)," (in Indonesian), *Jurnal Pendidikan dan Pranata Islam*, vol. 12, no. 1, pp. 41–51, 2021, doi: https://doi.org/10.36835/syaikhuna.v12i1.4363.
- [29] T. D. Cook and C. S. Reichardt, *Qualitative and quantitative methods in evaluation research*. Sage Publications, 1979, doi: 10.2307/1174260.
- [30] M. Q. Patton, "Two decades of developments in qualitative inquiry: a personal, experiential perspective," *Qualitative Social Work*, vol. 1, no. 3, pp. 261–283, Sep. 2002, doi: 10.1177/1473325002001003636.
- [31] C. Y. Piaw, *Asas statistik penyelidikan*, 3rd ed. Kuala Lumpur: Mc Graw Hill Education (in Malay), 2014.
- [32] M. Hammersley, "The issue of quality in qualitative research," *International Journal of Research and Method in Education*, vol. 30, no. 3, pp. 287–305, Nov. 2007, doi: 10.1080/17437270701614782.
- [33] C. Robson, *Real world research*, 3rd ed. Chichester: Wiley, 2011.
- [34] S. B. Merriam, *Qualitative research: A guide to design and implementation*, 3rd ed. San Francisco, CA: Jossey-Bass, 2009.
- [35] J. W. Creswell and C. N. Poth, *Qualitative inquiry and research design: Choosing among five approaches*. Sage Publications, 2016.
- [36] W. M. Trochim, *The research methods knowledge base*. Mason, Ohio: Cengage Learning, 2007.
- [37] O. Lebar, *Penyelidikan kualitatif: pengenalan kepada teori dan metode*. Perak: UPSI (in Malay), 2009.
- [38] C. Stenbacka, "Qualitative research requires quality concepts of its own," *Management Decision*, vol. 39, no. 7, pp. 551–556, Sep. 2001, doi: 10.1108/EUM000000005801.
- [39] R. K. Yin, *Case study research and applications: Design and methods*, 6th ed. Sage Publications, 2018.
- [40] M. Healy and C. Perry, "Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm," *Qualitative Market Research: An International Journal*, vol. 3, no. 3, 2000, doi: 10.1108/13522750010333861.
- [41] Y. S. Lincoln, E. G. Guba, and J. J. Pilotta, "Naturalistic inquiry," *International Journal of Intercultural Relations*, vol. 9, no. 4, pp. 438–439, Jan. 1985, doi: 10.1016/0147-1767(85)90062-8.

- [42] J. W. Creswell and D. L. Miller, "Determining validity in qualitative inquiry," *Theory into Practice*, vol. 39, no. 3, pp. 124–130, Aug. 2000, doi: 10.1207/s15430421tip3903\_2.
- [43] D. Davies and J. Dodd, "Qualitative research and the question of rigor," *Qualitative Health Research*, vol. 12, no. 2, pp. 279–289, Feb. 2002, doi: 10.1177/104973230201200211.
- [44] C. Seale, "Quality issues in qualitative inquiry," *Qualitative Social Work*, vol. 1, no. 1, pp. 97–110, Mar. 2002, doi: 10.1177/147332500200100107.
- [45] C. L. Kimberlin and A. G. Winterstein, "Validity and reliability of measurement instruments used in research," *American Journal of Health-System Pharmacy*, vol. 65, no. 23, pp. 2276–2284, Dec. 2008, doi: 10.2146/ajhp070364.
- [46] J. W. Creswell, *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications, 2017.
- [47] F. N. Kerlinger and H. B. Lee, *Foundations of behavioral research*. New York: Holt, Rinehart and Winston, 1986.
- [48] D. T. Campbell and D. W. Fiske, "Convergent and discriminant validation by the multitrait-multimethod matrix," *Psychological Bulletin*, vol. 56, no. 2, pp. 81–105, 1959, doi: 10.1037/h0046016.
- [49] J. R. Fraenkel, N. E. Wallen, and H. H. Hyun, *How to design and evaluate research*. USA: Mc Graw Hill Inc., 2007.
- [50] R. A. Lowe and S. B. Abbuhl, "Appropriate standards for 'appropriateness' research," *Annals of Emergency Medicine*, vol. 37, no. 6, pp. 629–632, Jun. 2001, doi: 10.1067/mem.2001.115216.
- [51] K. Bussey, "Lying and truthfulness: children's definitions, standards, and evaluative reactions," *Child Development*, vol. 63, no. 1, pp. 129–137, Feb. 1992, doi: 10.1111/j.1467-8624.1992.tb03601.x.
- [52] M. Young and J. Muller, "Truth and truthfulness in the sociology of educational knowledge," *Theory and Research in Education*, vol. 5, no. 2, pp. 173–201, Jul. 2007, doi: 10.1177/1477878507077732.
- [53] C. Kirkman, "Establishing truthfulness, consistency and transferability," *Nurse researcher*, vol. 15, no. 4, pp. 68–78, Jul. 2008, doi: 10.7748/nr2008.07.15.4.68.c6664.
- [54] V. H. Sutlive and D. A. Ulrich, "Interpreting statistical significance and meaningfulness in adapted physical activity research," *Adapted Physical Activity Quarterly*, vol. 15, no. 2, pp. 103–118, Apr. 1998, doi: 10.1123/apaq.15.2.103.
- [55] A. Chen, "Meaningfulness in physical education: A description of high school students' conceptions," *Journal of Teaching in Physical Education*, vol. 17, no. 3, pp. 285–306, Apr. 1998, doi: 10.1123/jtpe.17.3.285.
- [56] B. D. Baker, "Evaluating the reliability, validity, and usefulness of education cost studies," *Journal of Education Finance*, vol. 32, no. 2, pp. 170–201, 2006, [Online]. Available: <https://www.jstor.org/stable/40704290>.
- [57] V. Tseng and S. Nutley, "Building the infrastructure to improve the use and usefulness of research in education," in *Using Research Evidence in Education*, Springer International Publishing, 2014, pp. 163–175, doi: 10.1007/978-3-319-04690-7\_11.
- [58] J. W. Keefe, "Developing a defensible learning style paradigm," *Educational Leadership*, vol. 48, no. 2, pp. 57–61, 1979.
- [59] L. W. T. Schuwirth *et al.*, "When enough is enough: A conceptual basis for fair and defensible practice performance assessment," *Medical Education*, vol. 36, no. 10, pp. 925–930, Oct. 2002, doi: 10.1046/j.1365-2923.2002.01313.x.
- [60] D. R. Black, A. M. Taylor, and D. C. Coster, "Accuracy of self-reported body weight: stepped approach model component assessment," *Health Education Research*, vol. 13, no. 2, pp. 301–307, Jun. 1998, doi: 10.1093/her/13.2.301.
- [61] R. P. Leow, "Fluency and accuracy: toward balance in language teaching and learning," *Studies in Second Language Acquisition*, vol. 15, no. 2, pp. 267–269, Jun. 1993, doi: 10.1017/s0272263100012055.
- [62] N. D. McMillan *et al.*, "The instrumental engineering of a polymer fibre drop analyser for both quantitative and qualitative analysis with special reference to fingerprinting liquids," *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, vol. 114, pp. 75–97, Aug. 1996, doi: 10.1016/0927-7757(96)03537-6.
- [63] L. Cohen, L. Manion, and K. Morrison, *Research methods in education*. Routledge, 2007, doi: 10.4324/9780203029053.
- [64] S. R. Cotten, A. Tashakkori, and C. Teddlie, "Mixed methodology: combining qualitative and quantitative approaches," *Contemporary Sociology*, vol. 28, no. 6, p. 752, Nov. 1999, doi: 10.2307/2655606.
- [65] J. P. Spradley, *The ethnographic interview*. New York: Holt, Rinehart and Winston, 1979.
- [66] R. A. Howard, "Decision analysis: practice and promise," *Management Science*, vol. 34, no. 6, pp. 679–695, Jun. 1988, doi: 10.1287/mnsc.34.6.679.
- [67] Y. K. Djamba and W. L. Neuman, "Social research methods: qualitative and quantitative approaches," *Teaching Sociology*, vol. 30, no. 3, p. 380, Jul. 2002, doi: 10.2307/3211488.
- [68] R. C. Bogdan and S. K. Biklen, *Qualitative research for education an introduction to theory and methods*, 4th ed. New York: Pearson, 2003.
- [69] S. L. Gill, "Qualitative sampling methods," *Journal of Human Lactation*, vol. 36, no. 4, pp. 579–581, Aug. 2020, doi: 10.1177/08903344200949218.
- [70] A. Stone, K. Allen, T. R. Rhoads, T. J. Murphy, R. L. Shehab, and C. Saha, "The statistics concept inventory: A pilot study," in *Proceedings - Frontiers in Education Conference*, IEEE, 2003, p. T3D1-T3D6, doi: 10.1109/FIE.2003.1263336.
- [71] P. H. Anderson, L. Lawton, R. J. Rexeisen, and A. C. Hubbard, "Short-term study abroad and intercultural sensitivity: A pilot study," *International Journal of Intercultural Relations*, vol. 30, no. 4, pp. 457–469, Jul. 2006, doi: 10.1016/j.ijintrel.2005.10.004.
- [72] J. L. Huberty, M. Siahpush, A. Beighle, E. Fuhrmeister, P. Silva, and G. Welk, "Ready for recess: a pilot study to increase physical activity in elementary school children," *Journal of School Health*, vol. 81, no. 5, pp. 251–257, Apr. 2011, doi: 10.1111/j.1746-1561.2011.00591.x.
- [73] K. H. Morin, "Value of a pilot study," *Journal of Nursing Education*, vol. 52, no. 10, pp. 547–548, Oct. 2013, doi: 10.3928/01484834-20130920-10.
- [74] C. Houghton, D. Casey, D. Shaw, and K. Murphy, "Rigour in qualitative case-study research," *Nurse Researcher*, vol. 20, no. 4, pp. 12–17, Mar. 2013, doi: 10.7748/nr2013.03.20.4.12.e326.
- [75] B. Miles, Matthew, A. M. Huberman, and J. Sladana, *Qualitative data analysis: a methods sourcebook*. London: Sage, 2014.
- [76] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, Apr. 1960, doi: 10.1177/001316446002000104.
- [77] D. Griffie, "Research tips: Interview data collection," *Journal of Developmental Education*, vol. 28, no. 3, pp. 36–37, 2005.
- [78] H. P. Widodo, "Methodological considerations in interview data transcription," *International Journal of Innovation in English Language Teaching and Research*, vol. 3, no. 1, pp. 101–107, 2014.
- [79] L. Cohen, L. Manion, and K. Morrison, *Research methods in education*. Routledge, 2002, doi: 10.4324/9780203224342.
- [80] V. J. Anthony and G. M. Joanne, "Understanding interobserver agreement: The Kappa statistic," *Family Medicine*, vol. 37, no. 5, pp. 360–363, 2005.
- [81] C. Magnusson, G. Finnerty, and R. Pope, "Methodological triangulation in midwifery education research," *Nurse researcher*, vol. 12, no. 4, pp. 30–39, Apr. 2005, doi: 10.7748/nr2005.04.12.4.30.c5957.
- [82] N. L. Leech and A. J. Onwuegbuzie, "An array of qualitative data analysis tools: a call for data analysis triangulation," *School Psychology Quarterly*, vol. 22, no. 4, pp. 557–584, Dec. 2007, doi: 10.1037/1045-3830.22.4.557.




- [83] C. Scott, "Research on mail surveys," *Journal of the Royal Statistical Society. Series A (General)*, vol. 124, no. 2, p. 143, 1961, doi: 10.2307/2984129.
- [84] N. K. Denzin, *The research act: a theoretical introduction to sociological methods*. New York: Routledge, 2017, doi: 10.4324/9781315134543.
- [85] N. Mack, C. Woodsong, K. M. MacQueen, G. Guest, and E. Namey, *Qualitative research methods: a data collector's field guide*. USA: Family Health International, 2005.
- [86] S. J. Taylor, R. Bogdan, and M. L. DeVault, "Introduction to qualitative research methods: a guidebook and resource," *Revista Brasileira de Linguística Aplicada*, vol. 5, no. 1, pp. 1689–1699, 2016.
- [87] A. Alaszewski, *Using diaries for social research*. London: SAGE Publications Ltd, 2006, doi: 10.4135/9780857020215.
- [88] A. Henry and C. Thorsen, "Disaffection and agentic engagement: 'Redesigning' activities to enable authentic self-expression," *Language Teaching Research*, vol. 24, no. 4, pp. 456–475, Sep. 2020, doi: 10.1177/1362168818795976.
- [89] E. S. L. van Batenburg, R. J. Oostdam, A. J. S. van Gelderen, R. G. Fukkink, and N. H. de Jong, "Evaluating opportunities in Dutch EFL course books for developing pre-vocational learners' oral interactional ability," *Language Teaching Research*, vol. 24, no. 4, pp. 434–455, Oct. 2020, doi: 10.1177/1362168818804258.
- [90] S. Mohamed and K. A. Jasmi, "Teaching and Learning Preparation Practice among Islamic Education Preschool Teachers," *UMRAN - International Journal of Islamic and Civilizational Studies*, vol. 8, no. 2, pp. 67–77, Jun. 2021, doi: 10.11113/umran2021.8n2.459.
- [91] A. J. Spenader, P. M. Wesely, and C. Glynn, "When culture is content: Applications for content-based instruction in the world language classroom," *Language Teaching Research*, vol. 24, no. 4, pp. 476–495, Sep. 2020, doi: 10.1177/1362168818799768.
- [92] M. H. Rozali and K. A. Jasmi, "Practical teaching characteristics of Islamic education in secondary schools," *Journal of Pharmaceutical Negative Results*, vol. 13, pp. 5493–5499, 2022, doi: 10.47750/pnr.2022.13.S09.671.
- [93] J. Creswell, A. C. Klassen, V. P. Clark, and K. C. Smith, "Best practices for mixed methods research in the health sciences," 2011.

## BIOGRAPHIES OF AUTHORS






**Azwani Masuwai**    obtained her Bachelor's degree in Islamic Studies (Shari'ah Law) from Universiti Kebangsaan Malaysia, in 2003, and her Master's degree in Education (Pedagogy) from Universiti Pendidikan Sultan Idris, in 2017. Additionally, she holds diplomas in both Islamic Studies (Shari'ah) and Education and has 20 years of experience as a secondary school teacher. She is currently pursuing a Ph.D. in Islamic Education at Universiti Kebangsaan Malaysia. She has authored or co-authored over 20 academic articles and developed a module on cooperative learning. Her current research involves creating a self-assessment tool for teaching quality, focusing on self-assessment and teacher attributes for Islamic education. She can be reached via email at [azwanimasuwai@yahoo.com](mailto:azwanimasuwai@yahoo.com).



**Hafizhah Zulkifli**    born in Penang, Malaysia. She earned her Bachelor's and Master's degrees in Islamic Education from Universiti Kebangsaan Malaysia, in 2008 and 2011, respectively. She completed her Ph.D. in Islamic Education at the International Islamic University Malaysia (IIUM) in 2019. She is currently a lecturer at Universiti Kebangsaan Malaysia and has authored or co-authored over 30 academic articles. Her research interests include the philosophy of education for children, known as Hikmah (wisdom) pedagogy, as well as Moral and Islamic Education. Presently, she is focused on developing educational materials that are more aligned with Malaysian and Islamic cultural and traditional contexts. She can be reached via email at [hafizhah\\_zulkifli@ukm.edu.my](mailto:hafizhah_zulkifli@ukm.edu.my).



**Mohd Isa Hamzah**    earned his Bachelor's degree in Usuluddin from Al-Azhar University in 1992. He completed his Master of Letters and Ph.D. in Education at the University of Birmingham in 1995 and 2008, respectively. Since 2000, he has been serving as a senior lecturer at Universiti Kebangsaan Malaysia. His expertise lies in art and applied arts within Islamic Studies and curriculum development. He has authored or co-authored over 60 academic articles. Currently, his research focuses on developing educational models for Rohingya refugee children in Malaysia. He can be reached via email at [isa\\_hamzah@ukm.edu.my](mailto:isa_hamzah@ukm.edu.my).