EDUCATUM JSMT Vol. 12 SPECIAL ISSUE (2025)
ISSN 2289-7070 / e-ISSN 2462-2451(33-44)
https://ejournal.upsi.edu.my/index.php/EJSMT/index

Development of the Fuzzy Delphi Instrument for Identifying Elements of the PhyARECi-STEM Learning Module

Renol Afrizon[1,2], Lilia Ellany Mohtar[1*], and Mohd Syahriman Mohd Azmi[1]

[1]Department of Physics, Universiti Pendidikan Sultan Idris, Malaysia
[2]Department of Physics, Universitas Negeri Padang, Indonesia

**Corresponding author:* lilia@fsmt.upsi.edu.my

Received: 1 Jan 2025; **Revision:** 20 February 2025; **Accepted:** 25 March 2025; **Published:** 28 April 2025

To cite this article (APA): Afrizon, R., Mohtar, L. E., & Mohd Azmi, M. S. (2025). Development of the Fuzzy Delphi Instrument for Identifying Elements of the PhyARECi-STEM Learning Module. EDUCATUM Journal of Science, Mathematics and Technology, 12, 33-44. https://doi.org/10.37134/ejsmt.vol12.sp.4.2025

To link to this article: https://doi.org/10.37134/ejsmt.vol12.sp.4.2025

Abstract

This research aims to describe the process of developing the Fuzzy Delphi instrument to identify elements of the applied physics learning module integrated with STEM and Augmented Reality or AR (PhyARECi-STEM learning module). This questionnaire was developed based on the national standards of higher education analysis in Indonesia, Hilda Taba's curriculum model, module elements found in literature, interviews, and focus groups with experts. The study is descriptive. The 50-item questionnaire relates to learning outcomes, content, process, and assessment. Nine experts conducted the validity evidence of the Fuzzy Delphi questionnaire to ensure content validity. Subsequently, the questionnaire was distributed to 30 respondents to obtain reliability. The data analysis techniques used for validity and reliability were the Content Validity Index (CVI) and Cronbach's alpha. The study's results demonstrated the validity of the Fuzzy Delphi questionnaire, as evidenced by an I-CVI value within the range of 0.89 to 1, an S-CVI value of 0.99, and a Kappa value within the range of 0.89 to 1.00. Furthermore, the reliability of the questionnaire was also established, with a Cronbach alpha value of 0.974. Therefore, the Fuzzy Delphi instrument, validated and demonstrated to be reliable, is now ready to be employed to identify the fundamental elements of the PhyARECi-STEM learning module via expert consensus.

Keywords: Augmented Reality, Applied Physics, Fuzzy Delphi Instruments, Learning Module, STEM

INTRODUCTION

STEM education has provided many benefits for students in higher education and schools, such as improved learning outcomes, attitudes, creativity, and 21st-century skills [1], [2], [3], [4]. STEM education provides learning experiences for real-world problems by integrating multiple disciplines (physics, AR technology, and mathematics) and making engineering the center of STEM [5], [6]. Some examples of the application of STEM and AR that have been successfully implemented in physics education are AR technology used to enhance students' physics learning self-efficacy [7], AR technology in the PBL process to improve learning performance and attitude in physics education [8], AR was found to have increased students' interest in learning physics and creating a positive environment for developing an understanding of the concepts [9], improve the reasoning performance of STEM students in modern physics courses [10], and STEM PjBL method to improve students' self-efficacy for solving mechanics physics problems [11]. Therefore, one of the efforts made to support the implementation of STEM learning in higher education is to provide teaching materials needed by pre-service physics teachers.

Pre-service physics teachers still need knowledge, conception, and competence to implement STEM education and AR in applied physics courses. Pre-service physics teachers in Indonesia already have a good STEM perception, interest, and motivation to implement and develop STEM learning experiences [12]. However, lack of exposure to engineering design, context-free classroom learning activities, and prior knowledge of the subject matter to be studied [13], [14] have become problems experienced by pre-service physics teachers in implementing STEM. Therefore, pre-service physics teachers need teaching materials to improve the competencies required to implement STEM, such as learning modules.

A learning module is a complete and free learning unit focusing on achieving some stated learning objectives. A learning module is a unit that facilitates pre-service teachers in learning how to apply the concepts from the course [15]. In addition, learning modules can also be studied independently [16], [17]. The developing learning module needs to pay attention to various rules and procedures that must be adhered to and the appropriate elements so that the module that will be produced can genuinely provide the best effect. Module development is one way to build STEM education in higher education and improve student competency, critical thinking, active learning, spatial skills, and Islamic values [10]. The learning module uses elements based solely on users' needs, such as students with limited knowledge and experience [18]. In addition, learning module elements are only decided through one data source [19]. Thus, determining the elements of a learning module requires other views, such as expert views.

The Fuzzy Delphi method is a technique used to identify elements of learning modules based on expert consensus [20]. This method can be run by distributing the Fuzzy Delphi questionnaire to the respondent experts. The instrument's preparation is the first step using the Fuzzy Delphi method. The uniqueness of this method is the early involvement of experts in the instrument's preparation [21]. However, no Fuzzy Delphi questionnaire is available in a standard form, so researchers must first design the questionnaire before using it. This study aims to generate empirical data on the validity and reliability of the Fuzzy Delphi questionnaire to identify the elements contained in the learning module based on expert views. This questionnaire and empirical data are expected to be valuable tools for researchers and curriculum developers to understand better the elements that must be present in the learning module. The research questions that guide this study are as follows.

1. How valid is the Fuzzy Delphi Questionnaire used to identify the elements of the PhyARECi-STEM learning module?
2. How reliable is the Fuzzy Delphi Questionnaire used to identify the elements of the PhyARECi-STEM learning module?

METHODS

Sample

There are nine experts involved in content validation. The selection of nine experts was justified by their ability to provide critical feedback on the instrument being developed. Experts typically have at least 15 years of experience in the relevant field, ensuring that their insights significantly contribute to the instrument's content validity. The Fuzzy Delphi questionnaire was piloted on 30 expert respondents from universities in Indonesia. 20-30 people with characteristics similar to those of the research subjects are sufficient to meet the practical threshold for exploratory research or the pilot study [22], [23]. The sample selection technique used a purposive sampling technique with scientific consideration: expertise equivalent to the expertise required for the consensus of the module elements and at least five years of work experience in the field. The characteristics of the respondents can be seen in Table 1.

Table 1. Respondent characteristics

Criteria	Frequency	Percentage (%)
Gender		
Male	12	40
Female	18	60
Expertise		
Physics Curriculum/Pedagogy	3	10

Criteria	Frequency	Percentage (%)
STEM Education/Integrated Science Education	5	17
Instrumentation Physics/ Material Physics/ Biophysics/ Climatology/ Earth Physics	15	50
Learning Media/ Technology	4	13
Physics Learning Assessment	3	10
Educational Degree		
Master	25	83
Doctor of Philosophy	5	17
Working Experience		
No experience	0	0
Less than 5 years	0	0
5 – 15 years	26	86
16 – 25 years	2	7
More than 25 years	2	7
Structural Position		
Lecturer	29	97
Study Program Coordinator	1	3
Institution Location in Indonesian		
Sumatera	29	97
Borneo	1	3

Fuzzy Delphi Questionnaire Development

The fuzzy Delphi instrument began with a literature review search related to the elements of the learning module. In addition, interviews were conducted with nine experts on the four elements of the learning module as additional reference sources besides the literature review. Interviews were conducted face-to-face and offline. The results of the interviews were summarized in the form of transcripts, and the results were analyzed. Furthermore, the focus group discussion (FGD) technique was used. The FGD was conducted with seven experts from various fields to obtain expert views on forming the first round of the Fuzzy Delphi questionnaire instrument. This was done because all experts have backgrounds in different fields. The process of developing the Fuzzy Delphi Questionnaire in detail can be seen in Figure 1.

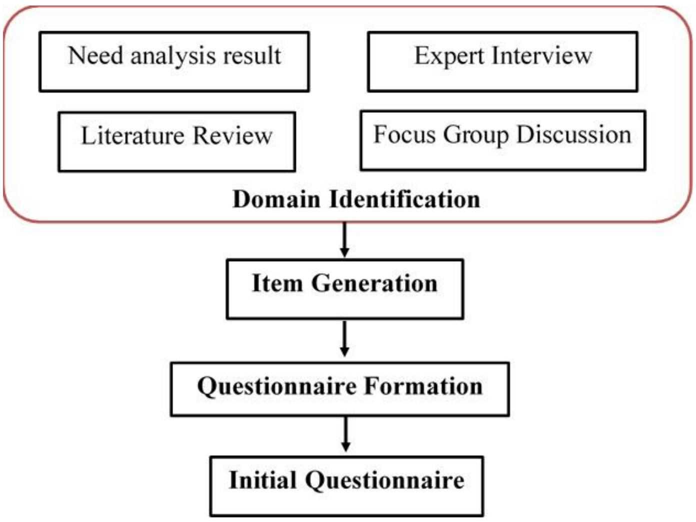

Figure 1. Fuzzy Delphi questionnaire development step

The FGD was conducted by gathering seven experts from various fields. The researcher invited a total of 9 experts, but only seven were able to attend the FGD. The results of the FGD in the transcript and theme are in accordance with the four main themes in the module's development: learning outcomes, learning content, learning process, and learning assessment. For learning outcomes, experts suggested that the items used be based on learning outcomes in the form of Program Learning Outcome (PLO), Course Learning Outcome (CLO), and sub-CLO of the physics education curriculum and the applied physics course syllabus. Then, for the content, all experts suggested using the content of the applied physics course curriculum using five topics. The topics of applied physics courses already in the syllabus are agricultural technology, alternative energy, digital technology, medical physics, global warming, and climate change. Experts suggest using the PhyARECi-STEM framework for the learning process. Finally, experts recommend using authentic, diagnostic, formative, and summative assessments for the evaluation theme.

Furthermore, the researcher created a fuzzy Delphi instrument of 50 items based on literature reviews, expert interviews, and findings from focused discussion analysis. According to expert recommendations, the Fuzzy Delphi instrument uses a 7-point linguistic scale: extremely disagree, very disagree, disagree, moderate agree, agree, very agree, and extremely agree. The 7-point scale is considered to provide the highest level of measurement validity and reliability when compared to other scales, including the 3-point, 5-point, and 10-point scales [24]. Moreover, employing a 7-point Likert scale in the FDM questionnaire has been demonstrated to minimize the ambiguity gap for each acceptance value of the expert agreement to a mere 3.3%, compared to a 5-point Likert scale, numbered 20% [25]. The researcher also used expert opinion by including additional open suggestions on each construct to obtain other expert views for the second round of fuzzy Delphi. The following is an example of a statement from the Fuzzy Delphi questionnaire in Table 2.

Table 2. Fuzzy Delphi questionnaire item example

Component	Example of Questionnaire Item
Learning Outcome	State your agreement regarding the learning outcomes that need to be in the applied physics learning module integrated with the STEM approach and Augmented Reality. e.g., Identifying examples of the application of physics in daily life
Learning Content	State your agreement regarding the learning content contained in the applied physics learning module integrated with the STEM approach and Augmented Reality. e.g., Application of mechanical concepts to agricultural technology and alternative energy
Learning Process	State your agreement regarding the learning process using the applied physics learning module integrated with the STEM approach and Augmented Reality so that pre-service physics teachers can use Augmented Reality as Technology Literacy. e.g., Visualizing physics concepts that are not directly visible (abstract)
Learning Assessment	State your agreement regarding the types of learning assessments that can be applied in the applied physics learning module integrated with the STEM approach and Augmented Reality. e.g., Authentic Assessment

Validity and Reliability in Questionnaire Development

Validity in developing a questionnaire at least includes the validity of the questionnaire's content. Content validity evidence must be done through expert agreement [26], [27], [28]. Several studies use the aspects of relevance and clarity as indicators that are often used to prove the validity of the content of the questionnaire [29], [30], [31]. The relevance aspect explains how related and directly useful an instrument is. Meanwhile, the clarity aspect shows how unambiguous an instrument is. Thus, content validity is defined as the degree of accuracy, truth, meaningfulness, and application of an instrument when measuring or observing what is to be measured or observed in terms of the relevance and clarity of the instrument. There are several stages in validating the contents of a questionnaire. First, prepare a content validation form. Second, select a review panel consisting of experts. Third, conduct content validation. Fourth, review domains and items. Fifth, give a score to each item. Sixth, calculate Content Validity Index (CVI) [32], [33]. Thus, the questionnaires can be validated, as the process can be seen in Figure 2.

Even though the questionnaire has been determined to be valid, it is still necessary to check whether the scales are reliable. Reliability refers to two situations, namely, the consistency of a measure and the stability of a measure over time [34], [35], [36]. Reliability estimation with the test-retest technique will produce a stability coefficient. Meanwhile, the alpha value and split-half method can be used to determine the instrument's internal consistency. Therefore, the reliability used in this study is to see the internal consistency of the questionnaire, where the process can be seen in Figure 2.

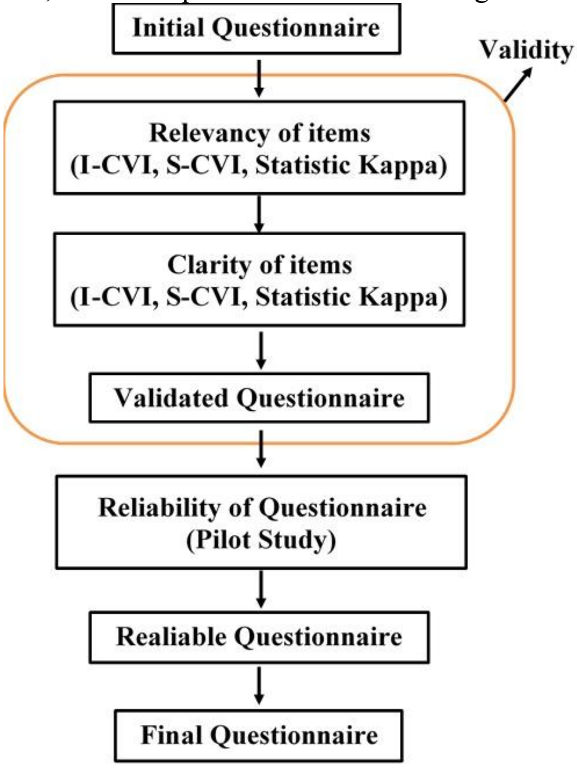

Figure 2. Validity and reliability of the Fuzzy Delphi questionnaire

Data analysis technique

The collected content validity data were analyzed using the CVI technique. The Content Validity Index presented in the form of Item-level Content Validity Index (I-CVI), Scale-level Content Validity Index (S–CVI), and Multi-rater Kappa Statistics (Modified Kappa) were used to prove the content validity of an instrument [37], [38], [39]. The range of I-CVI values is as follows. First, the item must be eliminated if I-CVI is less than 0.70. Second, if I-CVI is between 0.70 – 0.79, the statement item needs to be revised. Third, if the I-CVI value is more significant than 0.79, the statement item is appropriate and can be used [40]. In addition, the S-CVI value (average I-CVI) is more than 0.80, which is acceptable for indicating the validity of the instrument's content [40]. Multi-rater Kappa (Modified Kappa) can use the Equation 1 and Equation 2.

$$K = \frac{((I-CVI)-P_C)}{(1-P_C)} \qquad \textit{Equation 1}$$

$$P_C = \left[\frac{N!}{A!\cdot(N-A)!}\right] \times 0.5^N \qquad \textit{Equation 2}$$

Where Pc = N = Number of experts, A = Number of experts who choose a scale of 3-4 (score 1), Pc = probability of chance, K = Kappa statistical coefficient. The Kappa assessment criteria are as follows. First, a kappa value with a range of 0.80 - 1.00 has an almost perfect interpretation. Second, a kappa value with a range of 0.60 - 0.80 has a substantial interpretation. Third, a kappa value with a range of 0.40 - 0.60 has a sufficient interpretation. Fourth, a kappa value with a range of 0.20 - 0.40 has a moderate interpretation. Fifth, a kappa value of 0.00 - 0.20 has a less-than-ideal interpretation. Sixth, a kappa value with a value of 0.00 has an interpretation of no agreement. All experts agree on the results with an almost perfect interpretation [40]. Thus, data analysis using the CVI technique can be used to prove the validity of the content of an instrument.

Data analysis of an instrument's reliability uses the Cronbach alpha formula using Microsoft Excel or SPSS. The Cronbach alpha value is acceptable for high internal consistency. A small Cronbach alpha value of 0.70 indicates that the instrument is not reliable, while a minimum alpha value of 0.70 indicates that the instrument is reliable [42], [43], [44]. Therefore, the Cronbach alpha value can be used to prove the instrument's reliability.

RESULTS AND DISCUSSION

The research results are related to the proof of validity and reliability of the Fuzzy Delphi questionnaire. The results of the validity and reliability of the Fuzzy Delphi questionnaire are as follows.

Fuzzy Delphi Questionnaire Validity

Nine experts validated the formulated Fuzzy Delphi questionnaire to ensure that the items of each element of the questionnaire learning module were content-appropriate before they were distributed to experts for consensus. The results of the content validity evidence are in Table 3.

Table 3. The result of questionnaire content validity from relevancy and clarity aspect

Items	Relevant (rating 3 or 4)	Irrelevant (rating 1 or 2)	I-CVI	Pc	Kappa Statistic	Clarity (rating 3 or 4)	Clarity (rating 1 or 2)	I-CVI	Pc	Kappa Statistic
B.1	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
B.2	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
B.3	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
B.4	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
B.5	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
B.6	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
C.1.1	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
C.1.2	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
C.1.3	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
C.2.1	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
C.2.2	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
C.3.1	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
C.3.2	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
C.3.3	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
C.4.1	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
C.4.2	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
C.5.1	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
C.5.2	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
C.5.3	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
C.5.4	8	1	0.89	0.01563	0.89	8	1	0.89	0.01563	0.89
D.1.1	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
D.1.2	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
D.1.3	9	0	1.00	0.00195	1.00	8	1	0.89	0.01563	0.89
D.1.4	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
D.2.1	9	0	1.00	0.00195	1.00	8	1	0.89	0.01563	0.89
D.2.2.1	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
D.2.2.2	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
D.2.3.1	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
D.2.3.2	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
D.2.3.3	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
D.2.3.4	9	0	1.00	0.00195	1.00	8	1	0.89	0.01563	0.89
D.3.1	8	1	0.89	0.01563	0.89	8	1	0.89	0.01563	0.89
D.3.2	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
D.3.3	8	1	0.89	0.01563	0.89	8	1	0.89	0.01563	0.89
D.3.4	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
D.4.1	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
D.4.2	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
D.4.3	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
D.4.4	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
D.4.5	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
E.1.1.1	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
E.1.1.2	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
E.1.1.3	8	1	0.89	0.01563	0.89	9	0	1.00	0.00195	1.00
E.1.1.4	8	1	0.89	0.01563	0.89	9	0	1.00	0.00195	1.00
E.1.2.1	8	1	0.89	0.01563	0.89	9	0	1.00	0.00195	1.00

Items	Relevant (rating 3 or 4)	Irrelevant (rating 1 or 2)	I-CVI	Pc	Kappa Statistic	Clarity (rating 3 or 4)	Clarity (rating 1 or 2)	I-CVI	Pc	Kappa Statistic
E.1.2.2	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
E.1.2.3	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
E.2.1	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
E.2.2	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
E.2.3	9	0	1.00	0.00195	1.00	9	0	1.00	0.00195	1.00
Mean I-CVI =S-CVI			0.99					0.99		

Table 3 shows that the I-CVI value of each item of the Fuzzy Delphi questionnaire for the relevance aspect is in the range of 0.89 to 1. The S-CVI value for all items in the relevance aspect is 0.99. In addition, the I-CVI value for the clarity aspect is also in the range of 0.89 and 1. The S-CVI value for all items in the clarity aspect is also 0.99. The clarity and relevance of the questionnaire items provide an overview of the validity of the questionnaire's content [44]. The threshold value for content validity was obtained from 9 experts, whose I-CVI and S-CVI values are 0.78 and 0.80, respectively. The minimum CVI exceeds 0.75, indicating high suitability and acceptance by all experts [45]. Researchers must obtain a CVI value equivalent to 0.80 and above to prove the content validity of the new instrument [46]. A high I-CVI (usually above 0.78) indicates strong expert agreement that an item is relevant to the construct [47]. In addition, a high S-CVI (usually above 0.80) indicates that most items are considered relevant across the scale and ensures that the instrument is comprehensive in measuring the intended construct [47]. Conversely, certain items have been found to have a lower I-CVI score (0.89). While higher than the 0.78 threshold, this score means that some items are still valid. However, it is important to note that items with a lower I-CVI score have been revised based on expert feedback, as illustrated in Table 4.

Table 3 also shows each item's Kappa statistical (K) values regarding relevance and clarity. The items show K values above 0.74, namely 0.89 – 1.00 for both aspects, with almost perfect interpretation. A high K value (above 0.74 is considered excellent) indicates a high level of agreement between experts on the relevance of an item beyond what would be expected by chance. Kappa adds an extra layer by measuring how much agreement there is beyond chance [37]. On the other hand, certain items demonstrated a lower Kappa score (0.89). This score, while higher than the 0.74 threshold, indicates an almost perfect interpretation of the items in the questionnaire. However, items exhibiting lower Kappa scores underwent modifications, as evidenced in Table 4.

Table 4. Modified item statement based on expert feedback

Item	Initial Item Statement	Final Item Statement
C.5.4	How to overcome global warming/climate change	Global warming issues: impacts and solutions
D.1.3	Analyze socio-scientific issues and their benefits at the beginning of learning	Analyzing socio-scientific issues and the benefits at the beginning of learning
D.2.3.4	Representation of images, videos, animations, data, graphs, and mathematical equations with physics concept texts	Representing images, videos, animations, data, graphs, mathematical equations together with text
D.3.1	Ask, imagine, plan, create, and improve	Ask; imagine; plan; create; and improve
D.3.3	Define problems, plan possible solutions, select possible solutions, design, test, redesign, and communicate	Defining and researching problems; planning possible solutions; choosing the best solution; prototyping; testing; redesigning; and communicating

All items in the questionnaire are categorized as valid because they have I-CVI, S-CVI, and kappa statistical values that exceed the threshold. This value indicates that the Fuzzy Delphi questionnaire has good content validity. Similar research on the Fuzzy Delphi instrument for the Islamic cognitive domain also proved that the validity of the content with an S-CVI value of 0.80 (good content validity) involved relevant experts in the field with over a decade of experience [48]. Content validity is highly recommended before the questionnaire is used [49]. Thus, the Fuzzy Delphi questionnaire has been proven valid regarding content.

The questionnaire validation process was conducted by a panel of nine experts from various universities in Indonesia. Although the selection of experts was not based on probability sampling, several criteria were considered during the selection process. Firstly, the expert's expertise already represented the five expected areas of expertise. Secondly, the expert had worked in the field for at least 15 years. Thirdly, the experts were required to represent both internal and external organizations and be located on three Indonesian islands: Sumatra, Java, and Nusa Tenggara. The selection of experts was conducted in such a manner as to reduce the potential for bias [50]. Thus, the Fuzzy Delphi questionnaire, validated according to the opinions of the nine experts, was then used to determine the elements of the STEM learning module for pre-service physics teachers in Indonesia.

Fuzzy Delphi Questionnaire Reliability

After the instrument is validated, the reliability of the Fuzzy Delphi questionnaire needs to be carried out to check the internal consistency with the alpha formula if only one measurement is carried out. The results of the reliability proof can be seen in Table 5.

Table 5. Fuzzy Delphi questionnaire reliability results

N of Items	Cronbach's Alpha
50	0.974

From Table 5, all items have an alpha value of 0.974. All items have shown that the questionnaire has been internally consistent. The alpha value is more significant than 0.70, so the Fuzzy Delphi questionnaire is said to be reliable [51]. Cronbach's alpha is more than 0.90, indicating the questionnaire's consistency [52]. Items with a Cronbach's alpha value greater than 0.75 should be retained [53]. Research related to the Fuzzy Delphi instrument for the Islamic cognitive domain has proven that the Cronbach alpha value is 0.97, and researchers retain the items because their reliability is very high [48]. Therefore, the Fuzzy Delphi questionnaire has been reliable with a high Cronbach alpha value, so it can be used to identify elements of the PhyARECi-STEM learning module.

The Cronbach's Alpha value of 0.974 (> 0.90) is very high, which means that the instrument is potentially redundant [54]. At least 3 criteria to ensure redundancy in the instrument. First, analyze the correlation inter-item and whether the correlation value is very high (> 0.90) [54], [55], [56], [57], [58]. Second, analyze the deleted item to see if there is an increase in Cronbach's alpha value from the total alpha value of the items [54], [55]. Third, analyze the content of the item [58]. The findings show the inter-item correlation of the instrument ranged from 0.534 to 0.917, indicating that all items assessed closely related concepts. However, two items (C.3.1 and C.3.2) out of 50 had inter-item correlations above 0.90, which may indicate potential redundancy. In addition, Cronbach's alpha were calculated with each item removed, the Cronbach's alpha value would decrease slightly without drastically changing the items' overall alpha value. Furthermore, the content of item C.3.1 (introduction to the concept of electricity and magnetism) and item C.3.2 (application of the concept of electricity and magnetism in medical physics technology) were very different in terms of wording and meaning. Thus, all items did not meet the three criteria for redundancy of an instrument when assessed, so they were retained.

The reliability of the questionnaire involved 30 respondents from various universities in Indonesia. Although respondents were selected using non-probability sampling, several criteria were considered in selecting respondents in this study. First, the expertise of the respondents already represented the five expected expertise. Second, the experts have worked in the field for at least 5 years. Third, the experts have represented internal and external and come from 2 islands in Indonesia, namely Sumatra and Borneo. Based on these criteria, the potential for bias in selecting respondents can be avoided. Furthermore, purposive sampling techniques, which encompass a sufficient number of respondents, have been demonstrated to mitigate potential biases [59]. Thus, the findings show that the reliable Fuzzy Delphi questionnaire can be used to determine the elements of the STEM learning module for pre-service physics teachers in Indonesia.

CONCLUSION

The study results indicate that all items in the Fuzzy Delphi questionnaire have been proven valid and reliable. The current study provides an instrument that experts will use to consider and decide on the elements of a PhyARECi-STEM learning module. The results of this study can also be used as a reference source related to the feasibility of the Fuzzy Delphi instrument. The Fuzzy Delphi instrument, validated and proven reliable, is now ready to be employed to identify the fundamental elements of the PhyARECi-STEM learning module via expert consensus in Indonesia. Even though the sample selected in instrument validity evidence and the pilot study was adequate, it should be noted that it exclusively represented four Indonesian islands: Sumatra, Java, Borneo, and Nusa Tenggara. Consequently, there is an opportunity to conduct future research on the evidence of the instrument reliability in diverse educational contexts on other islands in Indonesia.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to 9 experts from universities in Indonesia who were willing to participate in the interview activities, FGD, and validate the Fuzzy Delphi questionnaire and 30 experts who were willing to be respondents whose data was used to determine the reliability of the instrument.

REFERENCES

[1] Amalya, C. P., Artika, W., Safrida, S., Nurmaliah, C., Muhibbuddin, M., & Syukri, M. (2021). Implementation of the problem base learning model combined with E-STEM based student worksheets on learning outcomes and self efficacy on environmental pollution materials. *Jurnal Penelitian Pendidikan IPA*, *7*(1), 37-38. https://doi.org/10.29303/jppipa.v7ispecialissue.962.

[2] Çeliker, H. D. (2020). The Effects of Scenario-Based STEM Project Design Process with Pre-Service Science Teachers: 21st Century Skills and Competencies, Integrative STEM Teaching Intentions and STEM Attitudes. *Journal of Educational Issues*, *6*(2), 451-477. https://doi.org/10.5296/jei.v6i2.17993.

[3] Chen, K., & Chen, C. (2021). Effects of STEM Inquiry Method on Learning Attitude and Creativity. *Eurasia Journal of Mathematics, Science and Technology Education*, *17*(11). https://doi.org/10.29333/EJMSTE/11254.

[4] Oschepkov, A. A., Kidinov, A. V., Babieva, N. S., Vrublevskiy, A. S., Egorova, E. V., & Zhdanov, S. P. (2022). STEM technology-based model helps create an educational environment for developing students' technical and creative thinking. *Eurasia Journal of Mathematics, Science and Technology Education*, *18*(5), em2110. https://doi.org/10.29333/ejmste/12033.

[5] Donmez, I. (2020). STEM education dimensions: From STEM literacy to STEM assessment. *Research Highlights in Education and Science 2020*, 154. https://files.eric.ed.gov/fulltext/ED611813.pdf#page=158.

[6] Roehrig, G. H., Dare, E. A., Ellis, J. A., & Ring-Whalen, E. (2021). Beyond the basics: A detailed conceptual framework of integrated STEM. *Disciplinary and Interdisciplinary Science Education Research*, *3*, 1-18. https://doi.org/10.1186/s43031-021-00041-y.

[7] Cai, S., Liu, C., Wang, T., Liu, E., & Liang, J. C. (2021). Effects of learning physics using Augmented Reality on students' self-efficacy and conceptions of learning. *British Journal of Educational Technology*, *52*(1), 235-251. https://doi.org/10.1111/bjet.13020.

[8] Fidan, M., & Tuncel, M. (2019). Integrating augmented reality into problem based learning: The effects on learning achievement and attitude in physics education. *Computers & Education*, *142*, 103635. https://doi.org/10.1016/j.compedu.2019.103635.

[9] Ropawandi, D., Halim, L., & Husnin, H. (2022). Augmented reality (AR) technology-based learning: the effect on physics learning during the covid-19 pandemic. *International Journal of Information and Education Technology*, *12*(2), 132-140. https://doi.org/10.18178/ijiet.2022.12.2.1596.

[10] Verawati, N., Ernita, N., & Prayogi, S. (2022). Enhancing the reasoning performance of STEM students in modern physics courses using virtual simulation in the LMS platform. *International Journal of Emerging Technologies in Learning (iJET)*, *17*(13), 267-277. https://doi.org/10.3991/ijet.v17i13.31459.

[11] Samsudın, M. A., Jamalı, S. M., Zaın, A. N. M., & Ebrahım, N. A. (2020). The effect of STEM project based learning on self-efficacy among high-school physics students. *Journal of Turkish Science Education*, *17*(1), 94-108. https://doi.org/10.2139/ssrn.3574024.

[12] Nugraha, M. G., Kidman, G., & Tan, H. (2023). Pre-service teacher in STEM education: An integrative review and mapping of the Indonesian research literature. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(5), em2262. https://doi.org/10.29333/ejmste/13155.

[13] Aisyah, N., Nawawi, E., & Yusup, M. (2020). How to support pre-service teachers' conceptual understanding using hyperthermia as STEM context: a case study in sciences education department Universitas Sriwijaya. In *Journal of Physics: Conference Series* (Vol. 1480, No. 1, p. 012072). IOP Publishing. https://doi.org/10.1088/1742-6596/1480/1/012072.

[14] Widiastuti, I., Budiyanto, C. W., & Ramli, M. (2020). Pre-service teacher conceptions of integrated STEM using the engineering design approach. In *2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)* (pp. 106-111). IEEE. https://doi.org/10.1109/TALE48869.2020.9368382.

[15] Houghton, J. (2023). Learning modules: problem-based learning, blended learning and flipping the classroom. *The Law Teacher*, *57*(3), 271-294. https://doi.org/10.1080/03069400.2023.2208017.

[16] Al Mamun, M. A., Lawrie, G., & Wright, T. (2020). Instructional design of scaffolded online learning modules for self-directed and inquiry-based learning environments. *Computers & Education*, *144*, 103695. https://doi.org/10.1016/j.compedu.2019.103695.

[17] Logan, R. M., Johnson, C. E., & Worsham, J. W. (2021). Development of an e-learning module to facilitate student learning and outcomes. *Teaching and Learning in Nursing*, *16*(2), 139-142. https://doi.org/10.1016/j.teln.2020.10.007.

[18] Malek, N. A., Abdullah, N. S. Y., Mat Darus, M., & Nursulistiyo, E. (2022). A need analysis for the development of Physics game-based interactive module in matriculation college. *EDUCATUM Journal of Science, Mathematics and Technology*, *9*, 48-60. https://doi.org/10.37134/ejsmt.vol9.sp.6.2022.

[19] Septiadevana, R., & Abdullah, N. (2024). Developing STEM project-based learning module for primary school teachers: a need analysis. *Int J Eval & Res Educ*, *13*(4), 2585-2593. https://doi.org/10.11591/ijere.v13i4.28894.

[20] Markus, L., & Ishak, M. Z. (2022). Development Of The Secondary School Quantum Physics-Stem (SSQP-Stem) Instructional Module For Physics Teachers: A Fuzzy Delphi Method (FDM) Approach. *International Research Journal of Education and Sciences (IRJES)*, *6* (2), 16–29. http://malrep.uum.edu.my/rep/Record/my.ums.eprints.38260.

[21] Al-Rikabi, Y. K., & Montazer, G. A. (2024). Designing an e-learning readiness assessment model for Iraqi universities employing Fuzzy Delphi Method. *Education and Information Technologies*, *29*(2), 2217-2257. https://doi.org/10.1007/s10639-023-11889-0.

[22] Karayağiz Muslu, G., Manav, G., Carti Dörterler, Ö., & Kara, R. (2024). Investigation of the psychometric properties of the Turkish child-friendly Oral Health Questionnaire. *BMC Oral Health*, *24*(1), 995. https://doi.org/10.1186/s12903-024-04764-y.

[23] Kudsi, Z., Fenlon, M. R., Hill, K., & Baysan, A. (2020). Developing a questionnaire to measure psychological disturbance associated with tooth loss. *Journal of Dentistry*, *98*, 103353. https://doi.org/10.1016/j.jdent.2020.103353.

[24] Ibrahim, M. Y. (2025). The most appropriate scale for behavioral research is a seven-point rating. *Multidisciplinary Reviews*, *8*(4), 2025126-2025126. https://doi.org/10.31893/multirev.2025126.

[25] Noh, N. M., Siraj, S., Halili, S. H., Jamil, M. R. M., & Husin, Z. (2019). Application of fuzzy delphi method as a vital element in technology as a tool in design thinking based learning. *Asia Pacific Journal of Educators and Education*, *34*, 129-151. https://doi.org/10.21315/apjee2019.34.7.

[26] López-Belmonte, J., Pozo-Sánchez, S., Lampropoulos, G., & Moreno-Guerrero, A. J. (2022). Design

and validation of a questionnaire for the evaluation of educational experiences in the metaverse in Spanish students (METAEDU). *Heliyon*, *8*(11). https://doi.org/10.1016/j.heliyon.2022.e11364.

[27] Silva, W. A. D., de Sampaio Brito, T. R., & Pereira, C. R. (2022). COVID-19 anxiety scale (CAS): Development and psychometric properties. *Current Psychology*, *41*(8), 5693-5702. https://doi.org/10.1007/s12144-020-01195-0.

[28] Elangovan, N., & Sundaravel, E. (2021). Method of preparing a document for survey instrument validation by experts. *MethodsX*, *8*, 101326. https://doi.org/10.1016/j.mex.2021.101326.

[29] Cocks, K., Wells, J. R., Johnson, C., Schmidt, H., Koller, M., Oerlemans, S., ... & European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group. (2023). Content validity of the EORTC quality of life questionnaire QLQ-C30 for use in cancer. *European journal of cancer*, *178*, 128-138. https://doi.org/10.1016/j.ejca.2022.10.026.

[30] Guillot-Valdés, M., Guillén-Riquelme, A., & Buela-Casal, G. (2022). Content validity through expert judgment for the Depression Clinical Evaluation Test. *International Journal of Clinical and Health Psychology*, *22*(2), 100292. https://doi.org/10.1016/j.ijchp.2022.100292.

[31] Roebianto, A., Savitri, S. I., Aulia, I., Suciyana, A., & Mubarokah, L. (2023). Content validity: Definition and procedure of content validation in psychological research. *TPM*, *30*(1), 5-18. https://doi.org/10.4473/TPM30.1.1.

[32] Ramly, S. N. F., Ahmad, N. J., & Yakob, N. (2022). Development, validity, and reliability of chemistry scientific creativity test for pre-university students. *International Journal of Science Education*, *44*(14), 1-16. https://doi.org/ 10.1080/09500693.2022.2116298.

[33] Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Education in medicine journal*, *11*(2), 49-54. https://doi.org/10.21315/eimj2019.11.2.6.

[34] Farkas, B. C., Krajcsi, A., Janacsek, K., & Nemeth, D. (2024). The complexity of measuring reliability in learning tasks: An illustration using the Alternating Serial Reaction Time Task. *Behavior Research Methods*, *56*(1), 301-317. https://doi.org/10.3758/s13428-022-02038-5.

[35] Vosloo, C., & Van Staden, W. (2024). Validity and reliability of the agentive steadfastness index. *International Journal of Psychology*, *59*(3), 378-389. https://doi.org/10.1002/ijop.13098.

[36] Zadorozhny, B. S., Petrides, K. V., Jongerling, J., Cuppello, S., & van der Linden, D. (2024). Assessing the temporal stability of a measure of trait emotional intelligence: Systematic review and empirical analysis. *Personality and Individual Differences*, *217*, 112467. https://doi.org/10.1016/j.paid.2023.112467.

[37] Handage, S., & Chander, M. (2021). Development of an instrument for measuring the student learning outcomes: A content validation process. *Indian Journal of Extension Education*, *57*(3), 1-7. https://doi.org/10.48165/ijee.2021.57302.

[38] Dalawi, I., Isa, M. R., Chen, X. W., Azhar, Z. I., & Aimran, N. (2023). Development of the Malay Language of understanding, attitude, practice and health literacy questionnaire on COVID-19 (MUAPHQ C-19): content validity & face validity analysis. *BMC Public Health*, *23*(1), 1131. https://doi.org/10.1186/s12889-023-16044-5.

[39] Gizaw, Z., Yalew, A. W., Bitew, B. D., Lee, J., & Bisesi, M. (2022). Development and validation of questionnaire to assess exposure of children to enteric infections in the rural northwest Ethiopia. *Scientific Reports*, *12*(1), 6740. https://doi.org/10.1038/s41598-022-10811-x.

[40] Roy, R., Sukumar, G. M., Philip, M., & Gopalakrishna, G. (2023). Face, content, criterion and construct validity assessment of a newly developed tool to assess and classify work–related stress (TAWS–16). *Plos one*, *18*(1), e0280189. https://doi.org/10.1371/journal.pone.0280189.

[41] Özkaya, G., Aydin, M. O., & Alper, Z. (2021). Distance education perception scale for medical students: a validity and reliability study. *BMC medical education*, *21*, 1-8. https://doi.org/10.1186/s12909-021-02839-w.

[42] Akbari-Chehrehbargh, Z., Sadat Tavafian, S., & Montazeri, A. (2020). The Back-care Behavior Assessment Questionnaire (BABAQ) for schoolchildren: development and psychometric evaluation. *BMC Public Health*, *20*, 1-10. https://doi.org/10.1186/s12889-020-09318-9.

[43] Kennedy, I. (2022). Sample size determination in test-retest and Cronbach alpha reliability estimates. *British Journal of Contemporary Education*, *2*(1), 17-29. https://doi.org/10.52589/bjce-fy266hk9.

[44] Kumar, A., Rajendran, A., Usman, M., Ahuja, J., Samad, S., Mittal, A., ... & Wig, N. (2022).

Development and validation of a questionnaire to evaluate the knowledge, attitude and practices regarding travel medicine amongst physicians in an apex tertiary hospital in Northern India. *Tropical Diseases, Travel Medicine and Vaccines*, *8*(1), 13. https://doi.org/10.1186/s40794-022-00170-w.

[45] Govindasamy, P., Cumming, T. M., & Abdullah, N. (2024). Validity and reliability of a needs analysis questionnaire for the development of a creativity module. *Journal of Research in Special Educational Needs*. https://doi.org/10.1111/1471-3802.12659.

[46] Masuwai, A., Zulkifli, H., & Hamzah, M. I. (2024). Evaluation of content validity and face validity of secondary school Islamic education teacher self-assessment instrument. *Cogent Education*, *11*(1), 2308410. https://doi.org/10.1080/2331186X.2024.2308410

[47] Shrotryia, V. K., & Dhanda, U. (2019). Content validity of assessment instrument for employee engagement. *Sage Open*, *9*(1). https://doi.org/10.1177/2158244018821751.

[48] Zulkifli, H., Ismail, A. M., Muhamad, N. A. F., Hussin, N. H., & Yusuf, S. A. M. (2024). Reliability and Validity of The Fuzzy Delphi Instrument of Islamic Cognitive Domains Running Head: Reliability and Validity of Islamic Cognitive Domains. *International Journal of Academic Research in Progressive Education and Development*, *13*(2). https://doi.org/10.6007/ijarped/v13-i2/21201.

[49] Taherdoost, H. (2016). Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research. *International Journal of Academic Research in Management (IJARM)*, *5*. https://doi.org/10.2139/ssrn.3205040.

[50] Von Soest, C. (2023). Why do we speak to experts? Reviving the strength of the expert interview method. *Perspectives on Politics*, *21*(1), 277-287. https://doi.org/10.1017/S1537592722001116.

[51] Sand, P., Kleiberg, A. N., Kljajić, M., & Lannering, B. (2020). The reliability of the health related quality of life questionnaire PedsQL 3.0 cancer module in a sample of Swedish children. *BMC pediatrics*, *20*, 1-8. https://doi.org/10.1186/s12887-020-02387-0.

[52] Zhang, J. Y., Wang, Y., Wan, B. J., Qi, L. H., Shu, T., & Feng, Z. C. (2023). Clinicians' response ability for public health emergencies: a questionnaire development and validation study. *Journal of Public Health*, 1-7. https://doi.org/10.1007/s10389-023-01951-z.

[53] Schmid, M., Brianza, E., & Petko, D. (2020). Developing a short assessment instrument for Technological Pedagogical Content Knowledge (TPACK. xs) and comparing the factor structure of an integrative and a transformative model. *Computers & Education*, *157*, 103967. https://doi.org/10.1016/j.compedu.2020.103967.

[54] Castillejos-Carrasco-Muñoz, R., Peinado-Rubia, A. B., Lérida-Ortega, M. Á., Ibáñez-Vera, A. J., Tapia-Toca, M. C., & Lomas-Vega, R. (2023). Validity and reliability of the Niigata PPPD Questionnaire in a Western population. *European Archives of Oto-Rhino-Laryngology*, *280*(12), 5267-5276. https://doi.org/10.1007/s00405-023-08038-1.

[55] Sims, J., Sloesen, B., Bentley, S., Naujoks, C., Arbuckle, R., Chiva-Razavi, S., ... & Wolffsohn, J. S. (2024). Psychometric evaluation of the near activity visual questionnaire presbyopia (NAVQ-P) and additional patient-reported outcome items. *Journal of Patient-Reported Outcomes*, *8*(1), 1-17. https://doi.org/10.1186/s41687-024-00717-9.

[56] Bom, J. A., Voormolen, D. C., Brouwer, W. B., de Bekker-Grob, E. W., & van Exel, J. (2024). Construct Validity, Reliability, and Responsiveness of the 10-Item Well-being Instrument for Use in Economic Evaluation Studies. *Value in Health*. https://doi.org/10.1016/j.jval.2024.02.014.

[57] Louis, E., Lee, W. J., Litcher-Kelly, L., Ollis, S., Pranschke, E., Fitzgerald, K., ... & Panés, J. (2024). Content Validity and Psychometric Evaluation of the Crohn's Symptom Severity (CSS) Questionnaire in Patients with Moderately to Severely Active Crohn's Disease. *Advances in Therapy*, *41*(9), 3678-3705. https://doi.org/10.1007/s12325-024-02923-9.

[58] Fokkens, W., Trigg, A., Lee, S. E., Chan, R. H., Diamant, Z., Hopkins, C., ... & Tabberer, M. (2023). Mepolizumab improvements in health-related quality of life and disease symptoms in a patient population with very severe chronic rhinosinusitis with nasal polyps: psychometric and efficacy analyses from the SYNAPSE study. *Journal of patient-reported outcomes*, *7*(1), 4. https://doi.org/10.1186/s41687-023-00543-5.

[59] López, M. (2023). The effect of sampling mode on response rate and bias in elite surveys," *Qual. Quant.*, 57 (2), 1303–1319. https://doi.org/10.1007/s11135-022-01406-9.