Pertanika J. Soc. Sci. & Hum. 34 (5): 2755 - 2775 (2026)

PERTANIKA JOURNALS

SOCIAL SCIENCES & HUMANITIES

Journal homepage: http://www.pertanika.upm.edu.my/

Content Validity and Internal Consistency of a Moon Phases Achievement Test and a Science Self-Efficacy Questionnaire for Year 5 Primary Science

Punnithann Subramaniam and Nor Asniza Ishak*

School of Educational Studies, Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia

ABSTRACT

This research examined the content validity as well as internal consistency of two instruments prepared for use in Year 5 primary science, the Moon Phases Achievement Test and the Science Self-Efficacy Questionnaire (SSEQ). The achievement assessment was designed by the researcher and contained 20 items covering the identification, sequence, and spatial and temporal relationships of Moon Phases. The 22-item self-efficacy questionnaire was adapted to measure vicarious experience, mastery experience, emotional state, as well as verbal persuasion. Each instrument underwent evaluation by an independent panel consisting of three experts. The relevance of each item was assessed using a four-point rating scale, while content validity was established through the calculation of the Item-Level Content Validity Index (I-CVI) and the Scale-Level Content Validity Index utilising the average approach (S-CVI/Ave). Every item demonstrated an I-CVI value of 1.00, and both instruments recorded an S-CVI/Ave value of 1.00. Expert comments nevertheless resulted in revisions to wording, visual presentation, item clarity, and construct representation before pilot testing. The revised instruments were administered to 60 Year 5 pupils from two schools. The achievement test recorded a KR-20 coefficient of 0.783 while the self-efficacy questionnaire demonstrated a Cronbach's alpha coefficient of 0.800. These results suggest that both instruments possess acceptable levels of content validity as well as internal consistency for the intended Year 5 primary science research context. Further psychometric evaluation is recommended before the instruments are applied more broadly.

ARTICLE INFO

Article history:
Received: 27 July 2026
Accepted: 28 September 2026
Published: 30 October 2026

DOI: https://doi.org/10.47836/pjssh.34.5.18

E-mail addresses:
punnithann@student.usm.my (Punnithann Subramaniam)
asnizaishak@usm.my (Nor Asniza Ishak)
* Corresponding author

Keywords: Content validity; internal consistency; Moon Phases Achievement Test; science self-efficacy; Year 5 primary science

INTRODUCTION

Achievement and science self-efficacy are important outcomes in science education.

© Universiti Putra Malaysia Press

Achievement reflects pupils' knowledge and understanding of scientific content, whereas science self-efficacy concerns their confidence in their capability to successfully undertake science-related activities. Prior research has indicated that self-efficacy is related to academic performance and science achievement (Bal-Taştan et al., 2018; Doménech-Betoret et al., 2017; Hasanah et al., 2021; Kirbulut & Uzuntiryaki-Kondakci, 2019). Therefore, accurate measurement of these outcomes requires instruments that are appropriate for the intended content, construct, and educational context.

This study focuses on two instruments for Year 5 primary science: a researcher-developed Moon Phases Achievement Test and an adapted Science Self-Efficacy Questionnaire (SSEQ). The Moon Phases topic is abstract and challenging for pupils (Chia et al., 2023; Hartanto et al., 2025; İnaltekin & Akbaba, 2024; Sağdıç, 2024), while science self-efficacy is relevant to how pupils' approach and perform science-related tasks (Bakar & Halim, 2022; Kirbulut & Uzuntiryaki-Kondakci, 2019; Ting et al., 2023). Since the two instruments have different purposes, structures and scoring systems, each instrument must be properly validated and tested for reliability prior to use in the larger study.

Achievement in the Moon Phases Topic

The Moon Phases topic is widely recognised as difficult for pupils to understand, given that it involves abstract astronomical concepts. Misconceptions include the belief that Moon Phases are caused by the Earth's shadow or clouds, and that the Moon produces its own light (LoPresto & Murrell, 2011; Türkmen, 2015). Difficulties in understanding the topic have also been reported in studies of astronomy learning and conceptual understanding (Chia et al., 2023; Syuhendri et al., 2022).

In the present study, a needs analysis drew on the 2023/2024 reports of *Pentaksiran Bilik Darjah* (PBD; classroom assessment) and *Ujian Akhir Sesi Akademik* (UASA; end-of-academic-session test). The reports indicated that Year 5 pupils attending *Sekolah Jenis Kebangsaan Tamil* (SJKT; Tamil national-type primary schools) in the Jempol district experienced difficulties answering questions on Moon Phases. Interviews with two science teachers also indicated difficulties in teaching the topic because of its abstract nature and limited support from concrete materials.

Previous research has developed instruments specifically for assessing understanding of lunar phases. Chastenay and Riopel (2020) developed the Moon Phases Concept Inventory for Middle School for pupils aged 10-14 and evaluated it with a sample that included Grade 5 pupils. Their study provides an established example of assessing lunar concepts among learners of a similar age to those in the present study. The measurement requirement in the present research concerns achievement in the Malaysian Year 5 Moon Phases curriculum, assessed through the particular content coverage, item formats and Tamil wording of the researcher-developed test.

The *Dokumen Standard Kurikulum dan Pentaksiran* (DSKP; Curriculum and Assessment Standard Document) for Year 5 Science provided the content basis for the achievement test, while the revised Bloom's taxonomy provided a framework for organising its cognitive demands (Anderson et al., 2001). The test addresses recognition of Moon Phases representations, names and sequence, and spatial and temporal relationships within a lunar cycle. These frameworks guided item development, but evidence was still needed concerning how well the resulting items represented the intended content. Sireci and Faulkner-Bond (2014) identify the definition, relevance and representation of the assessed domain as central considerations in evaluating test content. Accordingly, an expert review examined the suitability of the developed items, while pilot testing provided an initial estimate of the internal consistency of pupils' scores.

Science Self-Efficacy in Primary Science

Bandura's (1977) self-efficacy theory provides a conceptual basis for examining pupils' beliefs about their capability to undertake learning tasks. The theory identifies experiences of successful performance, observation of others, verbal persuasion and physiological states as sources of information from which individuals develop efficacy beliefs. In educational measurement, these sources are commonly represented as mastery experience, vicarious experience, social persuasion, and physiological and emotional states (Usher & Pajares, 2009). These sources concern the experiences and interpretations that inform capability beliefs; they should therefore be distinguished from direct judgements of capability.

Previous instrument-development studies illustrate how this distinction can guide measurement. Usher and Pajares (2009) developed a questionnaire assessing the four sources in middle-school mathematics and obtained evidence supporting a four-factor structure through successive stages of item development and evaluation. In primary science, Carroll et al. (2020) developed the Irish Science Self-Efficacy Children's Questionnaire (IS-SEC-Q) for investigating scientist-facilitated outreach among children aged 11-12. These studies provide a theoretical and empirical basis for examining sources of self-efficacy, while their findings relate to the instruments and populations they investigated.

The present 22-item SSEQ was adapted from the IS-SEC-Q and organised around mastery experience, vicarious experience, verbal persuasion and emotional state. Translation into Malay, contextual changes to the wording and modification from seven to five response categories produced a version requiring further evaluation. The International Test Commission (2017) recommends obtaining evidence for the intended use of adapted tests in the target population. The present study therefore examined the relevance and suitability of the adapted items through expert review and estimated internal consistency through pilot testing. The four proposed groups provided a

conceptual basis for this review; their factor structure was not tested in the present study.

The Need for Valid and Reliable Instruments

The purpose of instrument evaluation is to gather evidence supporting the interpretation of scores for an intended use. Sireci and Faulkner-Bond (2014) explain that evidence based on test content concerns the relationship between the assessment content and its stated purposes. For the Moon Phases Achievement Test, this required examination of the items against the specified curriculum content and cognitive demands. For the adapted SSEQ, it required examination of whether the wording remained relevant to the proposed sources of self-efficacy and appropriate for the intended pupils. Expert ratings and written comments provided complementary information about item relevance and areas requiring revision.

Internal consistency addresses a further measurement question: the extent to which item responses contribute consistently to the reported scores within the participating sample. However, an acceptable coefficient does not establish the dimensionality of a questionnaire (Taber, 2018). The present study therefore had two specific objectives for each instrument: to obtain content validity evidence through expert evaluation and to estimate internal consistency in a Year 5 pilot sample. KR-20 was used for the dichotomously scored achievement test, and Cronbach's alpha was used for the SSEQ. Together with the documented item revisions, these analyses provide preliminary evidence for the instrument versions prepared for the intended research context.

Research Gap and Rationale for the Study

The study arose from a need to prepare assessment instruments for a larger investigation of Year 5 primary science learning. The needs analysis, based on the 2023/2024 PBD and UASA reports and interviews with two science teachers, identified difficulties with the Moon Phases topic among the intended SJKT pupils. The achievement test was developed to assess identification of phase representations, knowledge of phase names and sequence, and understanding of spatial and temporal relationships within a lunar cycle. Its development therefore required decisions about curriculum coverage, cognitive demands, diagrams and wording in the Tamil translation. Examining these decisions was necessary to provide evidence that the test content corresponded to its intended assessment purpose (Sireci & Faulkner-Bond, 2014).

A related measurement requirement concerned the adapted SSEQ, which covered the four proposed sources of science self-efficacy. The questionnaire drew on the IS-SEC-Q developed by Carroll et al. (2020), but its use in the present context involved translation into Malay, changes to contextual wording and a reduction from seven to five response categories. Evidence concerning the original questionnaire could not, by itself, establish the suitability of this

modified version. The International Test Commission (2017) recommends obtaining reliability and validity evidence for adapted instruments in their intended populations. The present study addressed an initial part of this requirement through expert content review and pilot assessment of internal consistency.

The two instruments were intended to provide separate information about Moon Phases achievement and pupils' reported experiences relating to the sources of science self-efficacy within the same larger study. The present paper contributes preliminary measurement evidence for these instrument versions and documents the reasoning behind specific revisions. These include changes to astronomical diagrams and instructions in the achievement test, and more specific references to teachers, classmates and family members in the questionnaire. Reporting these decisions alongside item-level expert ratings allows readers to examine how the instruments were prepared and what was revised before pilot testing. The findings provide a basis for further evaluation of these versions within the intended Year 5 research context.

Research Objectives

This study aims to explore the validity and reliability of two instruments developed or adapted to be used in Year 5 primary science. In particular, the following goals were pursued:

1. To determine the content validity and internal consistency of the Year 5 Moon Phases Achievement Test.
2. To determine the content validity and internal consistency of the Science Self-Efficacy Questionnaire.

METHODOLOGY

Instrument Development, Adaptation and Preliminary Evaluation

This study examined the development of the Moon Phases Achievement Test and the adaptation of the Science Self-Efficacy Questionnaire (SSEQ), followed by content validation and pilot assessment of internal consistency. Both instruments were prepared for the same larger Year 5 science study, which required measures of achievement and pupils' experiences relating to sources of science self-efficacy. Their preparation is reported together to explain how these different measurement requirements were addressed. Each instrument had a separate evaluation objective, expert panel and internal consistency analysis.

The rationale for developing the achievement test was informed by a needs analysis involving the 2023/2024 Classroom Assessment (PBD) and Final Academic Session Test (UASA) reports for Year 5 SJKT pupils in the Jempol district, together with interviews with two science teachers. The reported difficulties provided a local basis for focusing on Moon Phases learning. The test was developed with reference to the Year 5 Science Curriculum and Assessment Standard Document (DSKP). Its intended abilities were identifying representations of Moon Phases, identifying and sequencing phase names, and interpreting spatial and temporal relationships within a lunar cycle.

A table of test specifications organised the content and cognitive demands using the revised Bloom's taxonomy (Anderson et al., 2001).

The SSEQ was adapted from the IS-SEC-Q developed by Carroll et al. (2020). Its preparation involved translation, contextual modification and a change from seven to five response categories. These changes required evaluation of the adapted version for its intended use, consistent with the International Test Commission (2017). Following development or adaptation and translation, both instruments underwent expert content review, revision based on written feedback and pilot administration to 60 Year 5 pupils. The evaluation reported here concerns content validity and internal consistency.

Year 5 Moon Phases Achievement Test

The Year 5 Moon Phases Achievement Test was developed with reference to the Year 5 Science DSKP (Ministry of Education, 2017). The test was designed to assess three main areas: i) identifying representations of Moon Phases, ii) mastering the names and sequence of the phases, and iii) understanding the spatial and temporal relationships involved in a complete lunar cycle.

The test contained 20 items: 16 multiple choices, 2 drawings, 1 matching, and 1 marking. The items were developed according to Bloom's updated taxonomy (Anderson et al., 2001). The difficulty distribution followed a ratio of 3:5:2, consisting of six low-difficulty items, ten moderate-difficulty items, and four high-difficulty items. Pupils were given one hour to complete the test. Correspondingly, a table of test specifications was prepared to ensure appropriate coverage of the intended content and cognitive levels. Table 1 presents the specification of the Year 5 Moon Phases Achievement Test.

The instrument assesses pupils' ability to i) identify drawings of the new and old crescent Moon, ii) identify and arrange the names and sequence of Moon phases, and iii) understand spatial and temporal relationships across one complete lunar cycle. Note that the achievement test was initially developed in Malay and subsequently translated into Tamil by an Excellent Science Teacher from SJKT. To ensure semantic equivalence, the Tamil version was then translated back into Malay by a Malay Language teacher from an SJKT following a back-translation procedure (Brislin, 1970; Creswell & Creswell, 2018; Gay et al., 2012). Subsequently, the back-translated Malay version was compared with the original Malay version to analyse discrepancies and ensure the intended meaning of each item was retained.

Science Self-Efficacy Questionnaire

Permission to use and adapt the IS-SEC-Q was obtained from the instrument developers prior to the adaptation. The questionnaire consisted of 22 items representing four dimensions of science self-efficacy: vicarious experience, mastery experience, emotional state, as well as verbal persuasion. Table 2 presents the constructs along with

Table 1
Specification of the Year 5 moon phases achievement test

Content Area	Remember	Understand	Apply	Analyse	Evaluate	Create	Total Items
Identifying representations of Moon Phases	1, 2, 3	4, 19, 20	–	–	–	–	6
Mastering the sequence and names of Moon Phases	–	–	7, 8, 9, 10, 11	5, 12, 14, 15, 16	–	–	10
Understanding spatial and temporal relationships in a complete lunar cycle	–	–	–	–	6, 13, 17	18	4
Total items	3	3	5	5	3	1	20
Difficulty level	Low	Low	Moderate	Moderate	High	High	
Difficulty distribution		6 low-difficulty items		10 moderate-difficulty items		4 high-difficulty items	3:5:2

Source: Authors' own work

Table 2
Constructs and number of items in the science self-efficacy questionnaire

Construct	Number of items
Mastery experience	6
Vicarious experience	4
Verbal persuasion	6
Emotional state	6
Total	22

Source: Authors' own work

the corresponding number of items included in the SSEQ.

The initial questionnaire employed a seven-point Likert scale. However, for the current study, the response format was adapted to a five-point Likert scale, with responses ranging from 1 (strongly disagree) to 5 (strongly agree). A five-point scale was selected as it provides a simpler response format for primary school pupils and reduces possible confusion in interpreting response options (Dawes, 2008; Revilla et al., 2014). The questionnaire contains five sections, namely i) Section A: demographic information, ii) Section B: mastery experience, iii) Section C: vicarious experience, iv) Section D: verbal persuasion, as well as v) Section E: emotional state.

Consequently, the questionnaire was translated into Malay and subjected to a back-to-back translation procedure. Two School Improvement Specialist Coaches Plus (SISC+) language officers with expertise in English language and Malay

language were involved in the translation process. The back-translated as well as translated versions were subsequently examined against the initial instrument to ensure consistency of meaning.

Content Validation

Separate panels of three experts evaluated the two instruments. The Moon Phases Achievement Test was reviewed by one university academic specialising in Science Education and two experienced science teachers. On the other hand, the SSEQ was reviewed by two university academics in education-related fields and one officer from the Science and Mathematics Unit of the Negeri Sembilan State Education Department. The experts were purposely selected based on their relevant academic qualifications, professional roles and expertise in Science Education, primary science teaching, STEM education, or science curriculum. The two school-based experts who evaluated the achievement test had over a decade of experience in the field of science education. Table 3 presents the profiles of the expert panels.

Both expert evaluation forms were reviewed by the research supervisor before they were distributed to the respective expert panels. Separate expert evaluation forms were used for the Moon Phases Achievement Test and the SSEQ. The achievement test evaluation form presented the study background, research objectives,

Table 3
Profiles of the expert panels

Instrument	Expert	Position	Institution	Field of Expertise/ Qualification
Year 5 Moon Phases Achievement Test	Expert 1	Professor	Universiti Putra Malaysia	PhD in Science Education
	Expert 2	Excellent Science Teacher, DG12	Sekolah Kebangsaan Nuri, Johol, Negeri Sembilan	B.Ed. in Primary Science Education
	Expert 3	Head of Science Panel, DG12	Sekolah Kebangsaan Tunku Kurshiah, Kuala Pilah, Negeri Sembilan	M.Ed. in Primary Science Education
Science Self-Efficacy Questionnaire	Expert 1	Senior Lecturer	Universiti Sains Malaysia	PhD in Biology Education and Educational Technology
	Expert 2	Senior Lecturer	Universiti Putra Malaysia	PhD in Science and STEM Education
	Expert 3	Assistant Director, Science and Mathematics Unit	Negeri Sembilan State Education Department	PhD in Primary Science Education

Source: Authors' own work

item formats, and table of test specifications. It identified the Year 5 Science Curriculum and Assessment Standard Document (DSKP) as the reference for curriculum coverage and the revised Bloom's taxonomy as the framework for classifying cognitive demands (Anderson et al., 2001). Experts were asked to consider the appropriateness of the format, language and terminology, curriculum coverage, item variety and intended difficulty levels. These materials provided a reference for examining the items against the stated assessment requirements.

The SSEQ evaluation form identified Carroll et al. (2020) as the source of the adapted questionnaire and presented definitions of mastery experience, vicarious experience, verbal persuasion and emotional state before their respective item groups. Experts were asked to consider item relevance, clarity, appropriateness of terminology and accuracy of meaning. The definitions and grouping of items provided a conceptual reference for examining whether the wording corresponded to the proposed sources of self-efficacy.

Both forms included instructions for assigning one relevance rating to each item using a four-point scale: 1 = not relevant, 2 = somewhat relevant, 3 = relevant and 4 = highly relevant. Space was provided for comments and suggested revisions. Providing explicit instructions, domain information and rating guidance is consistent with the preparation of content validation forms described by Yusoff (2019). Ratings of 3 or 4 were classified as indicating relevance when calculating the I-CVI, and the S-CVI/Ave was obtained by averaging the item-level indices (Polit & Beck, 2006). These indices concerned the items in the achievement test and SSEQ; they did not constitute a separate validation of the expert evaluation forms.

The I-CVI was determined as:

$$\text{I-CVI} = \frac{\text{Number of experts rating item as 3 or 4}}{\text{Total number of experts}}$$

The S-CVI/Ave was calculated as:

$$\text{S-CVI/Ave} = \frac{\text{Sum of all I-CVI values}}{\text{Total number of items}}$$

For panels consisting of three to five experts, an I-CVI of 1.00 was established as the required threshold for item acceptance, whereas an S-CVI/Ave of at least 0.80 was considered acceptable (Davis, 1992; Grant & Davis, 1997; Polit & Beck, 2006; Polit et al., 2007). Beyond the quantitative evaluations, the experts' written feedback, observations, and suggestions were reviewed and used to revise both instruments before pilot testing. Correspondingly, revised instruments were used in the pilot study.

Pilot Study

A pilot study was performed with 60 Year 5 pupils from two schools in Port Dickson. The schools were selected given that their pupils had age and background characteristics similar to those of the intended study population. Intact-group sampling was employed, whereby existing Year 5 classes were retained, and all pupils in the selected classes participated in the pilot study. No treatment or control groups

were formed, as the primary objective of the pilot study was to administer the same two instruments to all participants and to examine their reliability before use in the larger study.

The pupils were given one hour to complete the Moon Phases Achievement Test and 15 minutes to complete the SSEQ. The pilot study was implemented to assess the internal consistency of both research instruments as well as to determine whether further revisions were required before their use in the larger study.

Reliability Analysis

The Moon Phases Achievement Test used dichotomous scoring, with each response scored as correct (1) or incorrect (0). Although the test included multiple-choice, drawing, matching, and marking items, all responses were converted into dichotomous scores. Therefore, the KR-20 coefficient was applied to evaluate the instrument's internal consistency.

The SSEQ utilised a five-point Likert scale and was assessed through Cronbach's alpha. A coefficient of 0.70 or higher was regarded as an acceptable level of internal consistency, while values of 0.80 or above were interpreted as good (George & Mallery, 2003).

Ethical Considerations

Approval to conduct the pilot study and administer the research instruments was obtained from the Educational Planning and Policy Research Division, Ministry of Education (MoE) Malaysia [Approval No: KPM.600-3/2/3-eras (25668)]. The validation and pilot-testing procedures formed part of the larger research study and were conducted in accordance with the approved research procedures. Note that permission was obtained from the relevant authorities and participating schools before data collection.

Written informed consent was secured from the parents or legal guardians, while assent was obtained from the participating pupils prior to the commencement of data collection. Participation in the study was entirely voluntary, and all information gathered was utilised exclusively for research purposes. The identities of the participating pupils and schools were kept confidential in the reporting of the findings.

RESULTS

Content Validity of the Year 5 Moon Phases Achievement Test

Item-Level and Scale-Level Content Validity

Three experts evaluated the Year 5 Moon Phases Achievement Test, with all 20 items receiving ratings of 3 (relevant) or 4 (highly relevant) from each expert. Consequently, every item obtained an I-CVI of 1.00. The overall S-CVI/Ave was also 1.00, exceeding the minimum acceptable value of 0.80. Table 4 presents the item-level expert ratings as well as content validity indices.

Although all items met with the I-CVI criterion, the experts recommended several revisions to improve item presentation, wording, visual clarity and cognitive demand. Therefore, the quantitative CVI findings were considered together with the experts' written comments.

Table 4
Expert ratings and content validity indices for the Year 5 moon phases achievement test

Item	Expert 1	Expert 2	Expert 3	Agreement	I-CVI	Decision
1	4	4	4	3	1.00	Retained
2	4	4	4	3	1.00	Retained
3	3	4	4	3	1.00	Retained
4	3	4	4	3	1.00	Retained
5	4	4	4	3	1.00	Retained
6	3	3	4	3	1.00	Retained
7	4	4	4	3	1.00	Retained
8	4	4	4	3	1.00	Retained
9	4	4	4	3	1.00	Retained
10	4	4	4	3	1.00	Retained
11	4	4	4	3	1.00	Retained
12	4	3	3	3	1.00	Retained
13	4	3	4	3	1.00	Retained
14	4	4	4	3	1.00	Retained
15	4	4	4	3	1.00	Retained
16	4	4	4	3	1.00	Retained
17	3	4	4	3	1.00	Retained
18	4	4	4	3	1.00	Retained
19	4	4	4	3	1.00	Retained
20	4	4	4	3	1.00	Retained
S-CVI/Ave					1.00	Accepted

Source: Authors' own work

Expert Feedback and Revisions to the Achievement Test

The experts discovered the test's overall content was relevant to the Moon Phases topic but proposed several improvements. Consequently, these recommendations were incorporated before the instrument was used in the pilot study. Table 5 summarises the main feedback and revision actions.

The revisions mainly improved the clarity, visual representation, organisation and cognitive coverage of the test. No items were omitted as all 20 items met with the required I-CVI.

Content Validity of the Science Self-Efficacy Questionnaire

Item-Level and Scale-Level Content Validity

The 22-item SSEQ was evaluated by three experts according to its four constructs: vicarious experience, mastery experience, emotional state, as well as verbal persuasion. All items received ratings of 3 or 4 from all three experts and obtained an I-CVI of 1.00. The overall S-CVI/Ave was also 1.00. Table 6 presents the expert ratings according to the four constructs.

Table 5
Expert feedback and revisions to the Year 5 moon phases achievement test

Item	Expert Feedback	Revision Action
1, 2, 3, 4, 6, 7, 8, 10, and 11	Arrange answer options consistently from shorter to longer responses.	The answer options were rearranged from shorter to longer responses.
6	Revise the instructions to improve clarity.	The instruction was revised, and the answer options were adjusted accordingly.
12	Add the Sun to strengthen the representation of the Moon Phases concept.	A Sun diagram was added to the right side of the figure.
12	Add an orbital line to clarify the sequence of the Moon Phases.	A counterclockwise orbital line was added to the figure.
13	Improve the contrast between the illuminated and dark portions of the Moon.	The Moon diagram was revised using a darker colour setting.
13	Replace Roman numerals with alternative alphabetical labels.	Roman numerals were replaced with the letters K, L, M, and N.
18 and 20	Include items that require pupils to create or draw representations of the Moon Phases.	The items were revised as drawing tasks requiring pupils to produce Moon Phase representations.

Source: Authors' own work

Table 6
Expert ratings and content validity indices for the science self-efficacy questionnaire

Construct	Item	Expert 1	Expert 2	Expert 3	Agreement	I-CVI	Decision
Mastery experience	1	4	4	4	3	1.00	Retained
	2	4	4	4	3	1.00	Retained
	3	3	4	3	3	1.00	Retained
	4	4	4	4	3	1.00	Retained
	5	4	4	4	3	1.00	Retained
	6	3	4	4	3	1.00	Retained
Vicarious experience	7	3	4	4	3	1.00	Retained
	8	4	4	3	3	1.00	Retained
	9	3	4	4	3	1.00	Retained
	10	4	4	3	3	1.00	Retained
Verbal persuasion	11	4	4	4	3	1.00	Retained
	12	3	4	3	3	1.00	Retained
	13	4	4	4	3	1.00	Retained
	14	4	4	4	3	1.00	Retained
	15	4	3	4	3	1.00	Retained
	16	3	4	3	3	1.00	Retained
Emotional state	17	4	4	4	3	1.00	Retained
	18	4	4	4	3	1.00	Retained
	19	3	4	4	3	1.00	Retained
	20	4	3	3	3	1.00	Retained
	21	4	4	4	3	1.00	Retained
	22	3	4	4	3	1.00	Retained
S-CVI/Ave						1.00	Accepted

Source: Authors' own work

Although all items met with the required I-CVI value, several received ratings of 3, along with recommendations for wording, specificity, and construct representation. These comments were used to revise the questionnaire before pilot testing.

Expert Feedback and Revisions to the Science Self-Efficacy Questionnaire

The expert feedback mainly concerned the clarity and contextual appropriateness of the wording used for Year 5 pupils. Several general terms were replaced with more specific references, while selected items were rewarded to represent their respective self-efficacy constructs. Table 7 presents the main revision of actions.

The revisions included all 22 items, but increased linguistic clarity, specificity, age appropriateness, and consistency with the four intended sources of science self-efficacy.

Table 7
Expert feedback and revisions to the science self-efficacy questionnaire

Item	Expert Feedback	Revision Action
3	The original negatively worded item required recording and clearer wording.	Revised to: "I successfully complete difficult science tasks after working hard."
6	The term "most" was unnecessary.	The term was removed.
7	The term "adults" was too general.	"Adults" was replaced with "teacher."
8 and 10	The sentence structure was lengthy.	The wording was shortened by removing unnecessary phrasing.
9	The term "children" was too general.	"Children" was replaced with "classmates."
11, 12, 14, and 16	The phrase indicating that someone "told" the pupil did not clearly represent praise or encouragement.	The wording was revised to reflect praise more directly.
12	The term "people" was too general.	"People" was replaced with "teachers."
14	The term "good" required more appropriate wording.	The wording was revised using a more suitable expression of positive ability.
15	The original item reflected classmates' preference for collaboration rather than direct verbal persuasion.	Revised to reflect praise from classmates for the pupil's contribution to science tasks.
16	The term "adults" was too general.	"Adults" was replaced with "family members."
17	The original wording required a clearer expression of emotional tension.	Revised to: "I feel tense when doing science tasks."
18	The original wording required a clearer expression of fatigue.	Revised to: "Doing science tasks makes me feel tired."
19	The expression referring to feeling "stressed and nervous" was simplified.	The wording was revised to express the feeling "anxious."
20	The original item was considered more cognitive than emotional.	The item was revised to better reflect the emotional-state construct.
21	The sentence structure was revised for clarity.	Revised to: "I feel stressed when starting a science task."
22	The term referring to feeling "depressed" was considered unsuitable for the intended context.	Revised to reflect reduced interest when thinking about learning science.

Source: Authors' own work

Reliability of the Year 5 Moon Phases Achievement Test

The reliability of the Year 5 Moon Phases Achievement Test was analysed based on the pilot study of 60 Year 5 pupils. Each of the 20 items was rated dichotomously with 1 = correct and 0 = incorrect. The test obtained a KR-20 coefficient of 0.783, which is acceptable with regard to internal consistency.

Reliability of the Science Self-Efficacy Questionnaire

Cronbach's alpha was used to measure the internal consistency of the 22-item SSEQ. The Cronbach's alpha coefficient for the questionnaire was 0.800, which is considered to have good internal consistency. The results of reliability for both instruments are summarised in Table 8.

Overall, the results indicated that all items in both instruments had an I-CVI of 1.00, and the S-CVI/Ave of each instrument was 1.00, indicating that all items in both instruments met the specified content validity criterion. The expert feedback led to revisions in wording, visual presentation, item structure, and construct alignment, without removing any items. The subsequent pilot study produced a KR-20 coefficient of 0.783 for the achievement test and a Cronbach's alpha coefficient of 0.800 for the self-efficacy questionnaire.

DISCUSSION

This research investigated the content validity and internal consistency of two instruments intended for Year 5 primary science: the researcher-developed Moon Phases Achievement Test and the adapted SSEQ. The findings provide strong evidence of content validity for both instruments, while the pilot study showed acceptable to good internal consistency. The results are discussed in terms of the different purposes and characteristics of the two instruments.

Validity of the Year 5 Moon Phases Achievement Test

The Year 5 Moon Phases Achievement Test achieved an I-CVI of 1.00 for all 20 items and an S-CVI/Ave of 1.00. With three experts, an I-CVI of 1.00 indicates unanimous agreement regarding the relevance of every item, while the overall S-CVI/Ave exceeded the recommended minimum criterion (Davis, 1992; Polit & Beck, 2006; Polit et al., 2007). These findings provide evidence that the test content was considered appropriate for its intended purpose.

Table 8
Reliability results for the two instruments

Instrument	Number of Items	Reliability Analysis	Coefficient	Interpretation
Year 5 Moon Phases Achievement Test	20	KR-20	0.783	Acceptable
Science Self-Efficacy Questionnaire	22	Cronbach's alpha	0.800	Good

Source: Authors' own work

Other than that, the test's content validity was supported by its systematic construction. The items were aligned with the Year 5 Science curriculum and distributed across varying cognitive domains according to the updated Bloom's taxonomy framework (Anderson et al., 2001). This was important as the Moon Phases topic requires pupils to perform more than simple recall. Pupils must i) identify representations of the phases, ii) arrange them in sequence, and iii) understand spatial and temporal relationships across a complete lunar cycle. Previous studies have demonstrated that Moon Phases remain difficult for pupils and are associated with persistent misconceptions (Chia et al., 2023; Hartanto et al., 2025; İnaltekin & Akbaba, 2024; Syuhendri et al., 2022). Therefore, adequate coverage of the intended content and cognitive demands was necessary when constructing the achievement test.

Importantly, the perfect CVI values did not mean that no revisions were required. The experts recommended improvements to the organisation of answer options, item wording, diagram clarity, orbital representation, and the cognitive demand of selected items. These revisions demonstrate the value of considering both qualitative expert feedback and numerical CVI values. Expert judgement in instrument development should not be limited to determining whether an item is retained or rejected. However, it should also be used to improve content representation and item quality (Grant & Davis, 1997). Thus, the content-validation process strengthened the achievement test even though all items had already met the quantitative acceptance criterion.

Validity of the Science Self-Efficacy Questionnaire

The SSEQ likewise obtained an I-CVI of 1.00 for all 22 items and an S-CVI/Ave of 1.00. These results indicate complete agreement among the three experts on the relevance of each item to the intended questionnaire content (Polit & Beck, 2006; Polit et al., 2007).

In contrast to the achievement test, the questionnaire was modified based on the IS-SEC-Q introduced by Carroll et al. (2020). The original instrument provided the conceptual structure for measuring four dimensions of science self-efficacy, namely i) mastery experience, ii) vicarious experience, iii) verbal persuasion, as well as iv) emotional state. Nonetheless, adaptation involved translation, contextual modification, and a change from a seven-point to a five-point response scale. Hence, it was necessary to validate the adapted version.

Several items needed to be modified as indicated by the expert feedback, although they have been rated as acceptable. General terms like "adults", "children," and "people" were changed to more specific terms like "teacher", "classmates", "teachers," and "family members". Other items were amended for a more accurate reflection of verbal persuasion or emotional condition. These changes improved the clarity and context of the questionnaire for the target Year 5 primary school pupils.

Equivalence between the original and adapted versions was also achieved through the translation process. Back-translation is widely used to determine if the meaning of the translations is consistent with the meaning of the source instrument (Brislin, 1970). In this study, the translation and expert-review processes complemented one another by examining both linguistic equivalence and contextual suitability. The use concerning a five-point response scale was also considered more manageable for younger respondents, consistent with previous discussions concerning the number of response categories used in rating scales (Dawes, 2008; Revilla et al., 2014).

Nevertheless, the present findings should be clarified as evidence of content validity for the adapted questionnaire rather than comprehensive evidence of construct validity. Although the items were considered appropriate representations of the four proposed sources of science's self-efficacy, additional psychometric evaluation would still be necessary to investigate the latent factor structure of the adapted version.

Reliability of the Year 5 Moon Phases Achievement Test and Science Self-Efficacy Questionnaire

The Year 5 Moon Phases Achievement Test obtained a KR-20 coefficient of 0.783. This indicates acceptable internal consistency for the 20 dichotomously scored items (George & Mallery, 2003). The use of KR-20 was appropriate as all responses, including drawing, matching, and marking items, were ultimately scored as correct or incorrect.

The SSEQ obtained a Cronbach's alpha coefficient of 0.800, reflecting good internal consistency based on the interpretation adopted in this study (George & Mallery, 2003). This result suggests that the 22 items demonstrated satisfactory consistency in the pilot sample.

The use of different reliability coefficients reflected the different scoring characteristics of the two instruments. The achievement test assessed correct or incorrect performance, whereas the self-efficacy questionnaire measured responses on a Likert scale. Therefore, reliability coefficients should be interpreted according to the instrument's nature and intended use rather than treated as a single universal indicator of instrument quality (Taber, 2018).

The reliability findings support the use of both instruments in this intended study context. Nonetheless, internal consistency represents only one aspect of reliability. Hence, the coefficients reported in this study should be interpreted as evidence obtained from the pilot sample rather than as proof that the instruments will produce identical reliability estimates across all populations and settings.

CONCLUSION

This research investigated the reliability as well as validity of a researcher-developed Year 5 Moon Phases Achievement Test and an adapted SSEQ. Both instruments underwent systematic preparation, expert content validation, revision based on expert feedback, pilot testing, and reliability analysis.

All 20 items in the achievement test and all 22 items in the self-efficacy questionnaire achieved an I-CVI of 1.00, while both instruments obtained an S-CVI/Ave of 1.00. Note that the achievement test produced a KR-20 coefficient of 0.783, demonstrating an acceptable level of internal consistency. Meanwhile, the SSEQ recorded a Cronbach's alpha coefficient of 0.800, reflecting good internal consistency.

This study provides preliminary content validity and internal consistency evidence for two instruments prepared for a larger Year 5 science study. Its contribution lies in the evaluation of the developed and adapted versions and the documentation of item revisions arising from expert feedback. Further evaluation should examine achievement test item performance, the factor structure of the adapted SSEQ, and the interpretation of scores when the instruments are used to assess change following instruction.

IMPLICATION OF THE STUDY

Expert review was followed by revisions to both instruments and administration of the revised versions to 60 Year 5 pupils. Although all items met the specified relevance criterion, the written feedback identified changes needed in instructions, diagrams and questionnaire wording. All 20 achievement test items and 22 SSEQ items were retained. The subsequent pilot provided estimates of internal consistency for the revised versions, with KR-20 of 0.783 for the achievement test and Cronbach's alpha of 0.800 for the SSEQ. These stages documented the preparation and initial evaluation of the instruments for the larger study.

The intended next application was assessment before and after science instruction in the larger Year 5 study. The achievement test was prepared to assess performance on the specified Moon Phases content, while the SSEQ was prepared to collect pupils' reports concerning mastery experience, vicarious experience, verbal persuasion and emotional state in science learning. Their different purposes allowed the larger study to examine achievement and experiences associated with the proposed sources of self-efficacy. The present article reports the evidence obtained during instrument preparation; their subsequent application and any instructional outcomes are outside its scope.

The study's contribution includes the record of how expert feedback informed the preparation of these instrument versions. Tables 5 and 7 connect specific concerns to corresponding revisions, including changes to astronomical diagrams, instructions, social referents and expressions of encouragement. This record illustrates why unanimous relevance ratings can still be accompanied by recommendations for revision. It also allows readers to examine the reasoning behind the changes when considering the instruments for related research. Further use requires attention to the intended interpretation of scores and the evidence available for that purpose, consistent with Sireci and Faulkner-Bond (2014).

LIMITATIONS OF THE STUDY

This study possesses several limitations. Each instrument was evaluated by a separate panel of three experts, consistent with the minimum panel size discussed by Polit and Beck (2006). With three experts, an I-CVI of 1.00 requires all three to judge an item relevant by assigning a rating of 3 or 4. All items in the present study met this criterion. Consequently, the S-CVI/Ave was also 1.00 for each instrument because it was calculated by averaging the item-level indices. These findings document agreement within the participating panels. Nevertheless, the small number of experts limited the range of perspectives represented, and further review by a larger panel with varied expertise could provide additional insights into content coverage and item suitability.

The factor structure of the adapted SSEQ was not examined. Although expert review addressed item relevance and Cronbach's alpha provided an estimate of internal consistency, these findings do not establish whether the items form the four proposed groups. As Taber (2018) explains, an acceptable alpha coefficient does not demonstrate the dimensionality of a scale. This remains a limitation following the questionnaire's translation, contextual adaptation, and change in response format. Further research should examine the proposed structure through factor analysis. Sample planning for this work should consider the communalities of the items and the number of indicators representing each factor, as these influence sample-size requirements (MacCallum et al., 1999).

The pilot study involved 60 Year 5 pupils from two schools in Port Dickson. Bujang et al. (2024) propose a minimum of 30 respondents for preliminary questionnaire reliability assessment under the statistical assumptions considered in their review. The present sample exceeded this suggested minimum, providing support for its use at the pilot stage. However, meeting a general recommendation does not establish the precision of a particular reliability estimate; sample requirements also depend on the intended statistical power or precision (Bonett, 2002). Recruitment from intact classes in only two schools' further limits wider generalisation. The reported coefficients should therefore be interpreted as estimates obtained from this pilot sample.

Finally, the Moon Phases Achievement Test was developed specifically for the Year 5 Moon Phases topic and should not be interpreted as a general measure of primary science achievement. Similarly, the SSEQ was adapted for the intended study context, and further validation may be required before its use with different populations or settings.

ACKNOWLEDGEMENT

The authors would like to gratefully acknowledge the support and resources that Universiti Sains Malaysia has provided in the conduct of this research. Special thanks are due to the experts who evaluated the two instruments who gave valuable suggestions and constructive comments that have

contributed to the improvement of these two instruments.

REFERENCES

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (Complete ed.). Longman.

Bakar, M. R. A., & Halim, L. (2022). Pembinaan model literasi saintifik pelajar berdasarkan faktor persekitaran pembelajaran konstruktivisme, pengetahuan sains, sikap terhadap sains, efikasi kendiri dan motivasi. *Jurnal Pendidikan Malaysia, 47*(1), 1-18. https://doi.org/10.17576/JPEN-2022-47.01SI-01

Bal-Taştan, S., Davoudi, S. M. M., Masalimova, A. R., Bersanov, A. S., Kurbanov, R. A., Boiarchuk, A. V., & Pavlushin, A. A. (2018). The impacts of teacher's efficacy and motivation on student's academic achievement in science education among secondary and high school students. *EURASIA Journal of Mathematics, Science and Technology Education, 14*(6), 2353-2366. https://doi.org/10.29333/ejmste/89579

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215. https://doi.org/10.1037/0033-295X.84.2.191

Bonett, D. G. (2002). Sample size requirements for testing and estimating coefficient alpha. *Journal of Educational and Behavioral Statistics, 27*(4), 335-340. https://doi.org/10.3102/10769986027004335

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*(3), 185-216. https://doi.org/10.1177/135910457000100301

Bujang, M. A., Omar, E. D., Foo, D. H. P., & Hon, Y. K. (2024). Sample size determination for conducting a pilot study to assess reliability of a questionnaire. *Restorative Dentistry & Endodontics, 49*(1), Article e3. https://doi.org/10.5395/rde.2024.49.e3

Carroll, S., Sheahan, J., McCauley, V., & Grenon, M. (2020). *Development and validation of the Irish science self-efficacy children's questionnaire to assess the short-term influence of scientists facilitating outreach* [Preprint]. bioRxiv. https://doi.org/10.1101/2020.01.25.919357

Chastenay, P., & Riopel, M. (2020). Development and validation of the moon phases concept inventory for middle school. *Physical Review Physics Education Research, 16*(2), Article 020107. https://doi.org/10.1103/PhysRevPhysEducRes.16.020107

Chia, Y. K., Lim, S. H., Kiew, S. P., Koo, R. R. J., Lau, I. Z. W., Ling, Y. R., Chongo, S., & Lapawi, N. (2023). Overcoming students' misconception on moon phases: A classroom action research. *International Journal of Advanced Multidisciplinary Research and Studies, 3*(6), 375-379. https://www.multiresearchjournal.com/arclist/list-2023.3.6/id-1934

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE.

Davis, L. L. (1992). Instrument review: Getting the most from a panel of experts. *Applied Nursing Research, 5*(4), 194-197. https://doi.org/10.1016/S0897-1897(05)80008-4

Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research, 50*(1), 61-77. https://doi.org/10.1177/147078530805000106

Doménech-Betoret, F., Abellán-Roselló, L., & Gómez-Artiga, A. (2017). Self-efficacy, satisfaction, and academic achievement: The mediator role of students' expectancy-value beliefs. *Frontiers*

in Psychology, 8, Article 1193. https://doi.org/10.3389/fpsyg.2017.01193

Gay, L. R., Mills, G. E., & Airasian, P. W. (2012). *Educational research: Competencies for analysis and applications* (10th ed.). Pearson.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference* (4th ed.). Allyn & Bacon.

Grant, J. S., & Davis, L. L. (1997). Selection and use of content experts in instrument development. *Research in Nursing & Health, 20*(3), 269-274. https://doi.org/10.1002/(SICI)1098-240X(199706)20:3%3C269::AID-NUR9%3E3.0.CO;2-G

Hartanto, T. J., Mustika, M., & Budi, G. S. (2025). Improving prospective physics teachers' conceptual understanding of moon phases through an integrated observational and virtual simulation learning. *Prisma Sains: Jurnal Pengkajian Ilmu dan Pembelajaran Matematika dan IPA IKIP Mataram, 13*(4), 933-948. https://doi.org/10.33394/j-ps.v13i4.16338

Hasanah, R. S., Sholihin, H., & Nugraha, I. (2021). An investigation of junior high school students' science self-efficacy and its correlation with their science achievement in different school systems. *Journal of Science Learning, 4*(2), 192-202. https://doi.org/10.17509/jsl.v4i2.27275

İnaltekin, T., & Akbaba, B. (2024). Investigation of middle school students' model of astronomy events and information sources of incorrect model. *International Journal of Contemporary Educational Research, 11*(4), 508-524. https://doi.org/10.52380/ijcer.2024.11.4.680

International Test Commission. (2017). *The ITC guidelines for translating and adapting tests* (2nd ed.). https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf

Kirbulut, Z. D., & Uzuntiryaki-Kondakci, E. (2019). Examining the mediating effect of science self-efficacy on the relationship between metavariables and science achievement. *International Journal of Science Education, 41*(8), 995-1014. https://doi.org/10.1080/09500693.2019.1585594

LoPresto, M. C., & Murrell, S. R. (2011). An astronomical misconceptions survey. *Journal of College Science Teaching, 40*(5), 14. https://www.nsta.org/journal-college-science-teaching/journal-college-science-teaching-mayjune-2011/astronomical

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84-99. https://doi.org/10.1037/1082-989X.4.1.84

Ministry of Education. (2017). *Pelan pembangunan pendidikan Malaysia 2013-2025*. https://www.moe.gov.my/menumedia/media-cetak/penerbitan/dasar/1207-malaysia-education-blueprint-2013-2025/file

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29*(5), 489-497. https://doi.org/10.1002/nur.20147

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30*(4), 459-467. https://doi.org/10.1002/nur.20199

Revilla, M. A., Saris, W. E., & Krosnick, J. A. (2014). Choosing the number of categories in agree–disagree scales. *Sociological Methods & Research, 43*(1), 73-97. https://doi.org/10.1177/0049124113509605

Sağdıç, A. (2024). Shedding light on pre-service science teachers' instructional preferences and hurdles in teaching phases of the moon. *Kastamonu Education Journal, 32*(2), 201-214. https://doi.org/10.24106/kefdergi.1473470

Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema, 26*(1), 100-107. https://doi.org/10.7334/psicothema2013.256

Syuhendri, S., Siahaan, S. M., & Pasaribu, A. (2022). Lunar phases refutation texts: Supplement texts to overcome students' misconceptions. *European Journal of Education and Pedagogy, 3*(6), 157-160. https://doi.org/10.24018/ejedu.2022.3.6.504

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Ting, C. W., Surat, S., & Rahman, S. (2023). Tahap efikasi kendiri dan penguasaan kemahiran berfikir aras tinggi (KBAT) dalam kalangan murid. *Jurnal Dunia Pendidikan, 4*(4), 318-333. https://doi.org/10.55057/jdpd.2022.4.4.27

Türkmen, H. (2015). After almost half-century landing on the moon and still countering basic astronomy conceptions. *European Journal of Physics Education, 6*(2), 1-17. https://doi.org/10.20308/ejpe.70434

Usher, E. L., & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology, 34*(1), 89-101. https://doi.org/10.1016/j.cedpsych.2008.09.002

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