Full paper

Sains Humanika

# Content Validity and Reliability Analysis of Instruments on Students' Perceptions of Flipped STEM Learning

Lam Chau Ping, Nur Farha Shaafi*

*Faculty of Education and Sports Studies, Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia*

*Corresponding author: farhashaafi@ums.edu.my

**Article history**: Received: 05 December 2025 Received in revised form:30 January 2026 Accepted: 01 February 2026 Published online: 30 April 2026

**Abstract**

As the flipped classroom model continues to gain attention in secondary school STEM classrooms and educational quantitative research focusing on instrument development and validation, the need for a valid and reliable measurement instrument that is appropriate for the Malaysian secondary school context has become increasingly important. In STEM education, researchers seek to better understand students' learning experiences within flipped classroom practices within STEM-specific learning environments. Thus, this study aimed to investigate validity and reliability of the student perception of flipped classroom (SPFC) instrument; which measures; (i) collaborative environment, (ii) self-directed learning, (iii) learning impact, (iv) motivation and enjoyment, and (v) technology integration. Content validity was evaluated by ten experts through the application of the item-level content validity index (I-CVI), scale-level content validity index (S-CVI), and Kappa analysis. The study findings showed that the SPFC instrument demonstrated strong content validity (S-CVI/Ave = 0.96; S-CVI/UA = 0.85). Seventeen (17) items achieved full agreement (I-CVI = 1.00), while three items with lower CVI values were retained after clarity-focused revisions. A pilot study involving 30 students was conducted to assess reliability using Cronbach's alpha. The overall reliability was excellent ($\alpha$ = .927), with construct-level alphas ranging from acceptable to strong: collaborative environment ($\alpha$ = .866), self-directed learning ($\alpha$ = .685), learning impact ($\alpha$ = .873), motivation and enjoyment ($\alpha$ = .720), and technology integration ($\alpha$ = .884). In conclusion, the 20 items of SPFC questionnaire were accepted and retained to better understand students' perceptions of flipped classroom learning experiences in STEM context. Future research is recommended to further examine the instrument's construct validity and applicability across different STEM subjects and educational levels.

*Keywords*: Flipped classroom, STEM, CVI analysis, reliability

**Abstrak**

Seiring dengan peningkatan penggunaan model Flipped Classroom dalam amalan bilik darjah STEM sekolah menengah dan penyelidikan kuantitatif pendidikan yang memfokuskan pada pembangunan dan pengesahan instrumen, keperluan untuk instrumen pengukuran yang sah dan boleh dipercayai yang sesuai untuk konteks sekolah menengah Malaysia telah menjadi semakin penting. Dalam pendidikan STEM, penyelidik berusaha untuk lebih memahami pengalaman pembelajaran pelajar dalam amalan bilik darjah terbalik dalam persekitaran pembelajaran khusus STEM. Oleh itu, kajian ini bertujuan menyiasat kesahan dan kebolehpercayaan instrumen persepsi pelajar terhadap Flipped Classroom (SPFC), yang mengukur (i) persekitaran kolaboratif, (ii) pembelajaran kendiri, (iii) impak pembelajaran, (iv) motivasi dan keseronokan, serta (v) integrasi teknologi. Kesahan kandungan dinilai oleh sepuluh orang pakar melalui penggunaan indeks kesahan kandungan per item (I-CVI) dan indeks kesahan kandungan per skala (S-CVI) serta analisis Kappa. Dapatan kajian menunjukkan instrumen SPFC mencapai kesahan kandungan yang kukuh (S-CVI/Ave = 0.96; S-CVI/UA = 0.85). Tujuh belas (17) item mencapai persetujuan penuh (I-CVI = 1.00), manakala tiga item dengan nilai CVI yang lebih rendah dikekalkan selepas semakan. Kajian rintis yang melibatkan 30 orang pelajar telah dijalankan untuk menilai kebolehpercayaan menggunakan Cronbach alpha. Kebolehpercayaan keseluruhan adalah sangat baik ($\alpha$ = .927), dengan julat daripada boleh diterima sehingga kukuh: persekitaran kolaboratif ($\alpha$ = .866), pembelajaran kendiri ($\alpha$ = .685), impak pembelajaran ($\alpha$ = .873), motivasi dan keseronokan ($\alpha$ = .720), dan integrasi teknologi ($\alpha$ = .884). Kesimpulannya, 20 item soal selidik SPFC telah diterima dan dikekalkan untuk menilai persepsi pelajar terhadap pengalaman pembelajaran bilik darjah terbalik dalam konteks STEM. Penyelidikan masa depan disyorkan untuk mengkaji lebih lanjut kesahan dan kebolehgunaan konstruk instrumen merentas subjek STEM dan tahap pendidikan yang berbeza

*Kata kunci*: Flipped Classroom, STEM, analisis CVI, kebolehpercayaan



## 1.0 INTRODUCTION

The Sustainable Development Goals (SDGs), introduced by the United Nations in 2015 under the 2030 Agenda for Sustainable Development, emphasise the role of education and human capital development in addressing global challenges. In particular, SDG 4 (Quality Education) and SDG 9 (Industry, Innovation, and Infrastructure) highlight the need for strong competencies in science, technology, engineering, and mathematics (STEM) to support innovation-driven and knowledge-based economies (Jamaluddin et al., 2025; Milligan, 2022). STEM education is therefore not discussed in this study as a broad policy agenda, but as a foundational educational domain that equips learners with problem-solving, analytical, and technological skills essential for sustainable development.

Although its importance is widely acknowledged, STEM education continues to face longstanding challenges. Students often perceive STEM subjects as demanding, less enjoyable, and academically risky (Aspin et al., 2022; Ismail et al., 2019). Such perceptions reduce motivation and influence their willingness in pursuing science pathways (Fairhurst et al., 2023). In the Malaysian context, this issue is evident in enrolment trends. Recent statistics indicate that only 15.2% of Malaysian secondary school students enrolled in pure science subjects in 2024, with a marginal increase to 15.8% in 2025. Even after including another 5.22% students enrolled in STEM B (Science and Technology), overall participation remains worryingly low (Abdul Mutalib & Mohd Zulkanain, 2024; Hakim Mahari & Qistina Sallehuddin, 2025). These trends suggest declining interest and confidence in STEM learning, rather than a lack of awareness of its importance.

Prior research attributes this disengagement partly to instructional practices that remain predominantly teacher-centred, emphasising content delivery over active learning, problem solving, and collaboration (Nja et al., 2022). Such environments often limit active participation among students for meaningful interaction and self-directed learning, which are particularly critical in STEM education. While the COVID-19 pandemic exposed weaknesses in traditional and emergency remote teaching practices such as overdependence on passive teaching methods, insufficient digital competencies, and the absence of sustainable digital learning strategies (Divjak et al., 2022; Nolan et al., 2021), this study is positioned in a post-pandemic educational context. The relevance of prior pandemic-related findings lies not in the crisis itself, but in highlighting the ongoing need for sustainable, flexible, and learner-centred instructional approaches that remain effective beyond emergency conditions.

Given the consistent misalignment between the global importance of STEM and students' real classroom experiences, there is an urgent need for more innovative, student-centred instructional approaches. The flipped classroom model offers one such alternative. By shifting direct instruction outside of class and freeing classroom time for active engagement and problem solving, flipped learning aligns with UNESCO's vision for Education for Sustainable Development (2017) and has been shown to enhance student motivation, autonomy, and conceptual mastery, particularly in STEM contexts, where active problem solving is critical (Cho et al., 2021; Gong et al., 2024). Flipped classroom is not a new pedagogy but its innovation lies in its instructional design and implementation within specific contexts, particularly STEM classrooms where cognitive demand and learner support must be carefully balanced.

While existing literature extensively discuss the benefits, implementation, and general impact of flipped classrooms across various educational levels and disciplines including STEM, most studies focus on learning outcomes or engagement, with limited attention to directly address the specific challenge of validating measurement instruments for students' perceptions of flipped STEM learning in secondary schools in Malaysia. Recent studies suggested that the effectiveness of flipped instruction depends not merely on its adoption, but on how students experience its learning environment (Closs et al., 2021; Eren & Kenan Dikilitaş, 2025; Ma, 2023). Since students are the main beneficiaries of the instructional approach, it is important to have strongly validated instruments that can reliably capture students' perceptions.

The evaluation of the effectiveness of such pedagogical innovations from students' first point of view may help to provide a more meaningful and data-driven information to educators and policymakers in enabling improvements in pedagogical practices and curriculum design. Therefore, this study focuses on validating a questionnaire instrument to capture students' perceptions of the flipped classroom in STEM learning, ensuring both content validity and reliability. Specifically, the objective of the study is to determine the content validity and reliability of the questionnaire instrument on students' perception of the flipped classroom in STEM learning by using a detailed and systematic method, namely the content validity index (CVI) and Cronbach's alpha respectively.

## ■2.0 LITERATURE REVIEW

In the context of this research, the Flipped Classroom is defined as a blended learning model that combines both online and face-to-face instruction, where students are expected to engage with assigned pre-class materials such as videos, notes, or guided questions before participating in in-class active learning sessions. This approach was especially prominent during the post-COVID-19 recovery phase (Feijóo et al., 2021) as many teachers continued using digital platforms to distribute preparatory content before class. Such practice allows more equitable access to learning resources, particularly for students who struggled with traditional lecture-based instruction or who missed classes. Extensive international research has shown that the flipped classroom model can enhance student engagement (Agyeman & Aphane, 2024; Cheah & Sale, 2019; Karjanto & Acelajado, 2022; Sağlam & Arslan, 2018) self-directed learning (Aburayash, 2021; Chung & Lee, 2018; Öztürk & Çakıroğlu, 2021; Syakdiyah et al., 2018), and improve academic achievement across various subjects (Karadag & Keskin, 2017; Lo & Hew, 2019; Martínez-Jiménez & Ruiz-Jiménez, 2020; Talan & Gulsecen, 2019; Webb & Doman, 2020). These outcomes are particularly relevant to STEM education, where active participation and hands-on practice are essential for developing conceptual understanding. The model indeed offers significant benefits, including improved learning outcomes, higher student participation, and enhanced problem-solving skills (Kay & Dermott, 2018; Md Zainuddin et al., 2024).

Despite its benefits, the flipped classroom also presents challenges that may influence students' learning experiences. These include the need for reliable access to digital devices, increased out-of-class workload, and varying levels of student readiness or motivation to engage with pre-class materials (Pilu et al., 2025; Xiaoying & Abu Samah, 2024). All these factors that can collectively and negatively impact the overall effectiveness of the model, or worse, influence students' learning outcomes. Therefore, understanding how students perceive these aspects is extremely crucial, particularly when implementing flipped approaches in STEM subjects, which demand substantial conceptual engagement. Existing research on flipped learning has mainly focused on teachers, either pre-service or in-service (Ekinci et al., 2023; Kozikoğlu et al., 2021; Unal & Unal, 2021) or on its implementation outcomes (Elian & Hamaidi, 2018; Cho et al. 2021; Nja et al., 2022; Gong et al., 2024; Kalithas & Manivannan, 2024). There are relatively fewer studies that have used validated quantitative instruments to systematically measure students' perceptions, especially at the secondary school level, let alone in STEM context. Moreover, many studies rely on qualitative feedback alone, which is valuable but often subjective and context-specific, making broader comparison difficult (Bogaard et al., 2025).

This gap is critical because secondary school is a formative stage where students' STEM identities, career intentions, and attitudes solidify. Given that students are the primary beneficiaries of flipped instruction, their perceptions can offer actionable insights into what aspects of the model support or hinder learning.

The lack of validated instruments could limit the reliability, interpretive power, and generalizability of existing findings (Yusoff, 2019). This is because unvalidated tools risk measuring the constructs inconsistently, incompletely, or inaccurately (Karnia, 2024). As a result, educators and researchers risk relying on anecdotal or non-generalizable evidence that may not accurately and confidently capture students' experiences with flipped STEM instruction. To address this gap, the present study validates and assesses the reliability of a set of instruments designed to measure secondary students' perceptions of flipped STEM learning. Through expert review and reliability analysis, the study hopes to establish a validated and reliable measurement tool that supports evidence-based STEM pedagogy and enables future comparative research across diverse educational contexts. Establishing such a tool is also essential for advancing the broader goals of STEM education in relation to sustainability and the SDGs.

## ■3.0 METHODOLOGY

### 3.1 Research Design

This study employed a quantitative research approach using survey design. Survey-based designs are commonly used in the early phases of scale development to systematically evaluate item relevance, clarity, and internal consistency before progressing to construct validation (Creswell & Creswell, 2022; Stefana et al., 2025). The survey design was selected as it enables generalization from an accessible population comprising many individuals. Specifically, the design was used to facilitate the content validation and reliability testing of the draft scale. The primary purpose of this phase was not hypothesis testing or causal inference, but rather to establish preliminary psychometric soundness of the Students' Perception of Flipped Classroom (SPFC) instrument through: (1) expert-based content validation, and (2) pilot reliability analysis using student responses. It is recommended that at least three experts participate in the content validation process, but not more than ten (Bahari & Saleh, 2023; Lynn, 1986; Shrotryia & Dhanda, 2019). Content validity was examined using both Item-Level Content Validity Index (I-CVI) and Scale-Level Content Validity Index (S-CVI), consistent with established methodological recommendations (Lynn, 1986; Polit & Beck, 2006; Yusoff, 2019). For reliability assessment, a pilot test involving 30 students was administered using Cronbach's alpha, providing preliminary evidence of the scale's internal consistency and stability.

### 3.2 Research Sample

Content validation was conducted with a panel of ten experts selected through judgement sampling. This method was deemed appropriate because expert selection requires deliberate consideration of individuals' qualifications, scholarly contributions, and relevance to the study (Mat Said et al., 2022). Better validity was produced when mixed experts were chosen for both theoretical or research and real-world experiential knowledge (Intan & Yasin, 2021). The diversity of expertise strengthened the content validity of the scale and ensured that the instrument is both methodologically sound and responsive to the realities of secondary students' learning experiences in the flipped STEM classroom. The panel primarily comprised expertise in science education and educational technology, whereby, these domains constitute the core instructional foundations of STEM pedagogy at the secondary school level where STEM is operationalised through science instruction supported by technology-mediated learning environments (Rahman, 2024). This composition aligns with that of the scope of the study, focusing on students' learning experiences in flipped classroom implementations within science-based STEM instruction. Several experts also possessed explicit expertise in STEM education, curriculum design, and instructional innovation. Table 1 shows the summary of experts in this study.

**Table 1** Summary of the experts

| No. | Position | Fields of Expertise | Years of Experience |
|---|---|---|---|
| P1 | Senior Lecturer | Science Education, emerging technology in education, and game-based learning. | 20 |
| P2 | Professor | Science education | 31 |
| P3 | Professor | Science Education, STEM | 30 |
| P4 | School Senior Assistant | Chemistry Education | 21 |
| P5 | Senior Lecturer | Science Education, STEM | 19 |
| P6 | Physic Lecturer | Educational technology, science education | 10 |
| P7 | Senior Lecturer | Educational technology | 18 |
| P8 | Academic Teacher | Biology Education, educational technology | 21 |
| P9 | Senior Lecturer | Curriculum and instruction, science education | 20 |
| P10 | Senior Lecturer | STEM Education | 13 |

### 3.3 Instrumentation

The Students' Perception of Flipped Classroom (SPFC) instrument adapted the questionnaire originally developed by Johnson (2013) and later refined by Lam and Siew (2024). Several modifications were made to align the instrument with the objectives of this study, specifically to measure secondary students' perceptions of the Flipped Classroom instructions in STEM-oriented learning environments. Table 2 shows 20 items organised under five key dimensions in SPFC with respective codes.

Several items reflect behavioural tendencies, emotional responses, or learning preferences, whereby, these elements are conceptually aligned with students' holistic perceptions of their flipped classroom learning experiences, which inherently encompass cognitive, affective, and behavioural components (Bishop & Verleger, 2013). Thus, the construct is more accurately framed as students perceived flipped classroom learning experiences in a STEM context, rather than perception as a purely attitudinal construct.

**Table 2** Dimension and items for Students' Perception of Flipped Classroom (SPFC)

| Dimension | Code | Item Description |
|---|---|---|
| Collaborative environment | C1 | The Flipped Classroom is more engaging than traditional classroom instruction. |
| | C2 | The Flipped Classroom allows me to better communicate with other students. |
| | C3 | Flipped classroom gives me the opportunity to ask more questions inside the classroom. |
| Self-directed learning | S1 | I watch the video assignment regularly (at least twice a week). |
| | S2 | I dislike self-pace myself throughout the course. |
| | S3 | Learning STEM using Flipped Classroom strategy gave me the opportunity to self-learn. |
| | S4 | Frequently pause or repeat parts of the videos to increase my understanding of the material. |
| Learning impact | L1 | I feel that I understand the content of the STEM subjects better when I learn it through Flipped Classroom. |
| | L2 | The Flipped Classroom has not improved my learning of STEM. |
| | L3 | Preparing my pre-class material (notes, exercises, etc) according to Flipped Classroom strategy helped me to understand the lessons well. |
| | L4 | The Flipped Classroom strategy in learning STEM makes me exhausted. |
| | L5 | With Flipped Classroom model, I feel more prepared for my exams. |
| | L6 | Learning foundational content prior to class greatly enhances my understanding of material. |
| Motivation and enjoyment | M1 | I would not recommend Flipped Classroom to a friend. |
| | M2 | I am more motivated to learn STEM in Flipped Classroom. |
| | M3 | I feel fun in physical class meetings that implement the Flipped Classroom strategy. |
| | M4 | Flipped classroom improves my interest in exploring STEM topics. |
| Technology integration | T1 | I like watching the lessons taught on video. |
| | T2 | I would rather have a traditional fully teacher-led lesson than watching a video lesson. |
| | T3 | Learning STEM using Flipped Classroom strategy gave me the opportunity to engage with technology. |

The content validity questionnaire is implemented to measure the content validity of each dimension stated in Table 2. Guided by the rating scale proposed by Lynn (1986), the experts were required to tick only one response option for each item based on the scale from [1] not relevant; [2] somewhat relevant; [3] relevant but needs minor revision; and [4] very relevant for each item. This instrument contains 3 sections: section A: expert validation review form; section B: expert information; and section C: declaration of content validity review.

### 3.4 Content Validation

As this study represents an initial validation phase, content validity indices and adjusted kappa statistics were prioritized. Construct validity analyses such as Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) are planned for subsequent studies using larger samples, as recommended in scale development literature (Boateng et al., 2018). The validity of a research instrument can be defined as the capabilities of the questionnaire to measure a construct or variable to be measured (Creswell & Creswell, 2018). Meanwhile, content validity refers to the extent to which a measurement instrument accurately reflects what it is intended to measure (Polit & Beck, 2006; Yusoff, 2019; Yoong et al., 2022). There are two types of content validity indices (CVI): item-level content validity index (I-CVI), which evaluates the content validity of each item; and scale-level content validity index (S-CVI), which evaluates the content validity of the entire scale (Shi et al., 2012). Based on different calculation methods, S-CVI can be divided into two categories: S-CVI/Ave (Average Agreement) and S-CVI/UA (Universal Agreement). Table 3 summarises the definitions and formulas of the three content validity indexes (CVI) based on the recommendations by previous literature (Lynn, 1986; Polit et al., 2007; Polit & Beck, 2006; Yusoff, 2019) to be computed in Microsoft Excel to automate the validity calculation (Lukmanul Hakim et al., 2025).

**Table 3** Definitions and formulas of I-CVI, S-CVI/Ave and S-CVI/UA

| CVI | Definitions | Formulas |
|---|---|---|
| I-CVI | Proportion of experts rating an item as relevant (score 3 or 4) | $\frac{Number\ of\ experts\ rating\ item\ as\ 3\ or\ 4}{Total\ number\ of\ Experts}$ |
| S-CVI/Ave | Average of all I-CVI scores | $\frac{\sum \text{I}-\text{CVI}\ scores}{Total\ number\ of\ items}$ |
| S-CVI/UA | Proportion of items rated as relevant (score 3 or 4) by all experts. UA score will be 1 if and only if all experts agree (100% relevance), else UA score is 0. | $\frac{\sum \text{UA}\ scores}{Total\ number\ of\ items}$ |

Source: Lynn (1986); Polit & Beck (2006); Polit et al. (2007); Yusoff (2019), Lukmanul Hakim et al. (2025).

Due to the cost and time considerations, a non-face-to face approach was used where an online content validation form was sent to the experts along with clear instructions to ease the content validation process. During the expert consultation regarding content validity assessment, experts are requested to indicate the relevance or representativeness of each item to its respective content dimension. To prevent neutral responses and ensure choices are made, a four-point rating scale was used, with categories defined as follows: (1) not relevant, (2) somewhat relevant, (3) quite relevant, and (4) highly relevant. Through binary transformation, original scale rating of 1 and 2 are recoded as 0 while ratings of 3,4 will be 1 during CVI calculations as in Table 4.

**Table 4** Original scale and converted scale used for the CVI

| Original Scale | Converted Scale |
|---|---|
| 1, 2 | 0 |
| 3, 4 | 1 |

The acceptable CVI threshold varies according to the number of experts consulted. For two experts, the acceptable I-CVI was at least 0.80 (Davis, 1992). For three to five experts, it must be exactly 1.00 (Polit & Beck (2006); Polit et al., (2007)). For six to eight experts, it should be at least 0.83. For nine or more experts, a minimum value of 0.78 is required (Lynn, 1986). As ten experts were involved in evaluating the questionnaire in this study, the acceptable CVI threshold is therefore set to be 0.78 or higher. Table 5 summarizes the number of experts and the corresponding CVI thresholds. At the scale level, Polit and Beck (2006) recommend that S-CVI/Ave should not fall below 0.90, while Davis (1992) with Shrotryia and Dhanda (2019) suggested that S-CVI/UA should be no less than 0.80 to indicate strong content validity.

**Table 5** Number of experts and the corresponding acceptable content validity index (I-CVI) threshold

| Number of experts | Acceptable CVI values | Source |
|---|---|---|
| Two experts | ≥ 0.80 | Davis (1992) |
| Three to five experts | Exactly 1.00 | Polit & Beck (2006), Polit et al., (2007) |
| Six to eight experts | ≥ 0.83 | Lynn (1986) |
| At least nine experts | ≥ 0.78 | Lynn (1986) |

Shi et al. (2012) pointed that the consistent agreement among two or more experts on the items relevance might occasionally result from random selection rather than true consensus. Relying solely on the I-CVI can therefore overestimate validity because it does not account for random agreement. To address this, Polit et al. (2007) proposed calculating the probability of chance agreement (Pc) using the binomial distribution, assuming each expert randomly selects “relevant” or “not relevant” with equal probability (0.5). Pc represents the likelihood that the observed agreement occurred by chance, with formula:

$$P_c = \left[\frac{n!}{A!\,(n-A)!}\right] \times 0.5^n$$

where $n$ is the number of experts and $A$ is the number who rated the item as relevant (score 3 or 4). The adjusted kappa statistic (κ) is then derived by combining I-CVI and Pc to correct for chance agreement using:

$$\kappa = \frac{(I-CVI) - P_c}{1 - Pc}$$

This adjustment provides a more accurate measure of expert consensus beyond chance. Interpretation thresholds by Shi et al. (2012) and Rodrigues et al. (2017) for κ are: <0.40 (poor), 0.40–0.59 (fair), 0.60–0.74 (good), and ≥0.74 (excellent) are being used in this study (refer Table 6). Items with higher κ values indicate stronger content validity and are more likely to represent the intended construct. Including Pc and κ in the analysis ensures that retained items in the final instrument are based on genuine expert agreement, rather than random agreement.

**Table 6** Interpretation Guidelines for Adjusted Kappa (κ) Values (Rodrigues et al., 2017; Shi et al., 2012)

| Adjusted Kappa (κ) Values Range | Interpretation |
|---|---|
| κ ≥ 0.74 | Excellent |
| 0.60 ≤ κ < 0.74 | Good |
| 0.40 ≤ κ < 0.59 | Fair |
| κ < 0.40 | Poor |

### 3.5 Reliability Analysis

After the content validity analyses, reliability analysis follows as it plays an important role to ensure the usefulness of the test (Segal & Coolidge, 2018) by reflecting the stability and consistency of responses across repeated measurements (Creswell & Creswell, 2022). A pilot test involving 30 secondary students was conducted to examine the internal consistency of the SPFC instrument through Cronbach's alpha (α). In scale development research, pilot samples ranging from 20 to 40 participants are considered adequate for preliminary reliability estimation, particularly when the objective is to identify problematic items rather than to establish stable factor structures (Hertzog, 2008; Johanson & Brooks, 2010; Sukserm, 2024). The sample size was therefore deemed appropriate for this exploratory reliability phase, while recognising that larger samples will be required in subsequent studies for factor analysis and model confirmation.

Interpretation of reliability followed established psychometric guidelines. First, Nunnally and Bernstein (1994) recommend α ≥ 0.60 as acceptable for newly developed or adapted instruments, since early-stage scales may naturally display moderate internal consistency. Pallant (2020) later suggests that values approaching or exceeding α ≥ 0.70 represents good reliability for research applications, while values slightly below this threshold can still be adequate in preliminary or pilot studies. To provide a more detailed reliability classification, this study also referred the Cronbach’s alpha values guidelines proposed by George and Mallery (2019) as shown in Table 7.

**Table 7** Interpretation Guidelines of Cronbach alpha (α) values (George & Mallery, 2019)

| Cronbach Alpha (α) Values Range | Interpretation |
|---|---|
| $\alpha \geq 0.90$ | Excellent |
| $0.80 \leq \alpha < 0.90$ | Good |
| $.70 \leq \alpha < 0.80$ | Acceptable |
| $.60 \leq \alpha < 0.70$ | Questionable but usable |
| $0.50 \leq \alpha < 0.60$ | Poor |
| $\alpha < 0.50$ | Unacceptable |

## ■4.0 RESULTS

### 4.1 Content Validity Index (CVI) Analysis Scores

Table 8 presents the content validity evaluation for all 20 items measuring students' perceptions of flipped STEM learning. The analysis includes the item-level content validity index (I-CVI), universal agreement index (UA), probability of chance agreement (Pc), adjusted kappa (κ), expert comments, decisions and revisions made for each item if there is any. Majority of the items (17 out of 20) obtained perfect I-CVI of 1.00 with UA value of 1. Each of these items showed a low probability of chance agreement (Pc = 0.00098) together with a perfect adjusted kappa of 1.00. Two items (S2 and L4) recorded lower I-CVI values than the threshold of 0.78, obtaining only 0.60, and 0.70, respectively. A total of five items (C1, S1, S2, L4, T3) were revised based on expert feedback. Notably, several items (L4 and T3) with perfect consensus were revised as experts provided recommendations to refine wording and improve the precision of item statements despite their high relevance ratings. No items were removed from the scale. The remaining items were retained without changes. On the scale level, the overall instrument achieved an S-CVI/Ave of 0.96 and an S-CVI/UA of 0.85 based on the experts' evaluations.

At least two items (S2 and L4) recorded I-CVI values below the recommended threshold of 0.78, therefore, these items were revised based on both statistical evidence and qualitative expert judgment. Importantly, none of the experts rated these items as not relevant; instead, the lower agreement reflected concerns related to wording clarity and construct alignment, rather than the irrelevance of the underlying concept. In early-stage instrument development, items may be retained for revision when expert feedback indicates conceptual importance but recommends refinement to improve interpretability (DeVellis, 2016; Boateng et al., 2018; Lamm et al., 2020).

The adjusted kappa value (κ = 0.50) for item S2 indicated fair agreement, therefore, the expert comments consistently suggested grammatical and phrasing issues rather than construct redundancy. Following revision, the item was retained to preserve coverage of students' affective responses toward self-paced learning, which forms an integral component of perceived self-directed learning experiences in flipped classrooms. Similarly, item L4 was revised following expert consensus that the original wording overlapped with emotional fatigue or cognitive load rather than learning impact. The revised version reframed the item to reflect higher-order cognitive engagement, thereby strengthening alignment with the learning impact construct. The acceptable adjusted kappa value (κ = 0.66) further indicates that expert agreement exceeded chance levels, supporting its retention after revision. As this study represents an initial validation phase, re-validation after item revision was not conducted at this stage. Instead, revised items are intended to be subjected to further validation, including factor analytic procedures in subsequent future larger scale studies.

**Table 8** Content Validity Evaluation for SPFC Items (I-CVI, UA, Pc, κ, Comments, Decision, New item, S-CVI/Ave and S-CVI/UA)

| Item | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | No. of Agreement | I-CVI | UA | Pc | κ | Comments | Decision | New item after revision |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 8 | 0.80 | 0 | 0.04395 | 0.79 | 2/10 panels recommended simplifying the item to ensure better understanding by the respondents. | Revised | *The Flipped Classroom allows better in-class engagement with peers compared to traditional classroom lesson* |
| C2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 1.00 | 1 | 0.00098 | 1.00 | | Accepted | |
| C3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 1.00 | 1 | 0.00098 | 1.00 | | Accepted | |
| S1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 1.00 | 1 | 0.00098 | 1.00 | 1/10 panels suggested changing 'twice a week' to 'regularly as instructed' to be sound more wholesome for respondent. | Revised | *I watch the video assignment regularly as instructed.* |
| S2 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 6 | 0.60 | 0 | 0.20508 | 0.50 | 4/10 panels commented that the item needs to be revised for its grammar construction. Item is not deleted as none of the panels deemed it as not relevant. | Revised | *I dislike pacing my own learning throughout the course.* |
| S3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 1.00 | 1 | 0.00098 | 1.00 | | Accepted | |
| S4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 1.00 | 1 | 0.00098 | 1.00 | | Accepted | |
| L1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 1.00 | 1 | 0.00098 | 1.00 | | Accepted | |
| L2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 1.00 | 1 | 0.00098 | 1.00 | | Accepted | |
| L3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 1.00 | 1 | 0.00098 | 1.00 | | Accepted | |
| L4 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 7 | 0.70 | 0 | 0.11719 | 0.66 | 3/10 panels commented that the item measures more on cognitive load or emotion instead of learning impact. Panels suggested reframing if retained | Revised | *I spend more time analyzing and evaluating information when Flipped Classroom strategy is used in learning STEM.* |
| L5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 1.00 | 1 | 0.00098 | 1.00 | | Accepted | |
| L6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 1.00 | 1 | 0.00098 | 1.00 | | Accepted | |
| M1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 1.00 | 1 | 0.00098 | 1.00 | | Accepted | |
| M2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 1.00 | 1 | 0.00098 | 1.00 | | Accepted | |
| M3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 1.00 | 1 | 0.00098 | 1.00 | | Accepted | |
| M4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 1.00 | 1 | 0.00098 | 1.00 | | Accepted | |
| T1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 1.00 | 1 | 0.00098 | 1.00 | | Accepted | |
| T2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 1.00 | 1 | 0.00098 | 1.00 | | Accepted | |
| T3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 | 1.00 | 1 | 0.00098 | 1.00 | 1/10 panels suggested the item need to be more specific as the use of just 'technology' is too general to address the effectiveness of the integration. | Revised | *My ability to apply technology in academic projects has improved due to Flipped Classroom implementation in STEM learning* |
| | | | | | | | | | | | **S-CVI/Ave** | | 0.96 | | | | | |
| | | | | | | | | | | | **S-CVI/UA** | | 0.85 | | | | | |

### 4.2 Reliability analysis scores

Table 9 displays the Cronbach's alpha values for each construct in the SPFC, ATSM and CLQ instruments as well as their overall score. The overall scale demonstrated an excellent Cronbach's alpha of 0.927. The alpha values for the constructs' items ranged from 0.685 to 0.884. Three constructs had good Cronbach alpha with Technology Integration showing the highest value of 0.884, followed by Learning Impact (0.873) and Collaborative Environment (0.866). Motivation and Enjoyment with 0.720 is at an acceptable level while Self-Directed Learning recorded 0.685, despite questionable but still usable.

**Table 9** The Cronbach's alpha value of each construct items for SPFC

| Constructs | Cronbach's Alpha Value | Interpretation |
|---|---|---|
| Collaborative environment | 0.866 | Good |
| Self-directed learning | 0.685 | Questionable but usable |
| Learning impact | 0.873 | Good |
| Motivation and enjoyment | 0.720 | Acceptable |
| Technology integration | 0.884 | Good |
| **Overall** | 0.927 | Excellent |

## ■5.0 DISCUSSION AND RECOMMENDATION

The purpose of this study was to validate and assess the reliability of an instrument designed to measure secondary students' perceptions of the flipped classroom in STEM learning. The findings provide strong evidence that the instrument demonstrates both conceptual relevance and internal consistency, supporting its suitability for use in school-based and research-based contexts. The high content validity of the SPFC instrument as portrayed by high S-CVI/Ave=0.96 (≥0.90), S-CVI/UA=0.85 (≥0.80) and strong I-CVI values for most items (≥ 0.78), indicated strong content representativeness. These values exceeded the thresholds recommended by Lynn (1986), Polit & Beck (2006), Davis (1992) with Shrotryia and Dhanda (2019), demonstrating that the items collectively capture the essential dimensions of flipped STEM learning. Although certain items recorded lower I-CVI values compared to others, these items were retained since expert comments indicated that the issues were related to wording instead of conceptual misalignment. This suggests that even when initial agreement on item wording varies, expert consensus still affirms the importance of the construct being measured. The acceptable adjusted kappa values, $\kappa = 0.50$ (fair) $\kappa = 0.79$ (good) demonstrated by both items (S2 and L4) further confirm that the agreement was not due to chance. Thus, supporting their relevance and justifying their retention following revision. The revisions made to other three items despite their high CVI scores were not arbitrary. Instead, the amendments were made to enhance clarity, remove ambiguity, and ensure stronger alignment with the intended objective without compromising conceptual integrity. This reflects a commitment to methodological rigor and the iterative nature of instrument development, where statistical evidence should be balanced with practical considerations of readability and respondent comprehension.

Reliability analysis supported the internal consistency of the instrument. Overall Cronbach's alpha value demonstrated excellent reliability, while construct-level alphas ranged from acceptable to good. This implies that the instrument functions cohesively as a measure of students' flipped classroom experiences. At the construct level, four out of five constructs achieved acceptable to good reliability, with Collaborative Environment, Learning Impact, and Technology Integration demonstrating particularly strong consistency. This indicates that the items within each dimension cohesively measure students' perceived experiences of flipped classroom learning. This supports the theoretical assumption that key pedagogical features commonly associated with flipped learning such as peer interaction, pre-class preparation, and technology-mediated instruction can be meaningfully captured as structured perception-based constructs when operationalized through a well-designed instrument. The Self-Directed Learning construct produced the lowest alpha value ($\alpha = 0.685$), which is slightly below the conventional 0.70 benchmark but acceptable for early-stage development or pilot studies, especially when the goal is to refine and improve the instrument (Nunnally & Bernstein, 1994). This finding may reflect the inherent complexity of measuring self-regulatory behaviours among secondary students, who vary widely in their readiness for autonomous learning. It is also possible that the inclusion of a negatively worded item (S2) influenced response patterns, as negatively phrased items can introduce cognitive load and increase measurement error. While the construct remains meaningful, future refinement may improve internal consistency, such as rephrasing or replacing items that contribute to inconsistency.

To make up for the limitations presented in this study, future research may consider employing several rounds of expert validation to strengthen agreement on item relevance and wording. Next, expanding the diversity of field experts involved in the validation process to strengthen the breadth and representativeness of expert judgments, particularly for constructs that showed lower initial agreement. Engaging experts from various STEM disciplines and institutions or school settings such as international or private schools would enrich the quality of feedback and support greater generalizability. Additionally, preparing a larger and more varied item pool during the early development stage could improve reliability, especially for constructs such as Self-Directed Learning, where response patterns may be influenced by item phrasing or construct complexity. A broader item pool allows for more rigorous item selection and facilitates the removal or refinement of weaker items without compromising construct representation.

In addition, the incorporation of other advanced psychometric testing such as Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and measurement invariance testing across student groups and instructional contexts could be considered for future studies to further refine and validate the SPFC instrument. Such enhancements would strengthen its robustness and generalizability. Hence, enabling its application across diverse educational levels, including higher education institutions and supporting the development of other rigorously validated research instruments in flipped and STEM-oriented learning environments.

## ■6.0 CONCLUSION

In conclusion, the Students' Perception of Flipped Classroom (SPFC) instrument demonstrates strong content validity and acceptable internal consistency based on expert evaluation and pilot testing. The use of item-level and scale-level content validity indices provided quantitative evidence that the instrument items are relevant, essential, and representative of the intended dimensions. The findings indicated that students' perceived flipped classroom learning experiences in STEM-oriented science instruction are multidimensional, encompassing collaborative engagement, self-directed learning tendencies, perceived learning impact, motivation, and technology integration. These dimensions are coherently reflected in the structure of the SPFC instrument. However, it is important to note that this study represents an initial phase of instrument development. While content validity and reliability have been established, further validation through construct validity testing such as EFA and CFA is required before the instrument can be claimed as fully validated for large-scale or comparative research. In short, the SPFC instrument is intended to be administered as a self-report questionnaire to secondary school students who have experienced a structured flipped classroom implementation, where instructional content is delivered prior to class through video-based or digital materials and in-class time is devoted to active learning activities. Within these instructional conditions, the instrument provides a credible and practical tool for examining students' perceived learning experiences and informing improvements in flipped STEM instructional design. This may provide more meaningful insights for educators and researchers seeking to better understand and improve flipped learning environments.

### Acknowledgement

The researchers would like to thank the Universiti Malaysia Sabah for awarding the Cluster Fund Research Grant (DKP) with Research Code: DKP0105 to fund this publication and to the Ministry of Higher Education (KPT) Malaysia for funding the Early Career Fundamental Research Grant Scheme (FRGS) for 2024, Grant No. FRGS/1/2024/SSI09/UMS/02/7.

### Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper

## References

Abdul Mutalib, E., & Mohd Zulkanain, M. (2024). *Sorotan kemerosotan mata pelajaran Sains dan Matematik: Fenomena biasa pendidikan negara?* Astroawani.com. https://www.astroawani.com/berita-malaysia/kolumnis-sorotan-kemerosotan-mata-pelajaran-sains-dan-matematik-fenomena-biasa-pendidikan-negara-475930 retrieved on 23 June 2024

Aburayash, H. (2021). The students attitudes' toward the flipped classroom strategy and relationship to self-learning skills. *Journal of Education and Learning (EduLearn)*, 15(3). DOI: https://doi.org/10.11591/edulearn.v15i3.18132

Agyeman, N. Y. B., & Aphane, V. (2024). Implementing Flipped Classroom to Enhance Student Engagement: An Action Research. *E-Journal of Humanities, Arts and Social Sciences*, 5(11), 1860–1878. DOI: https://doi.org/10.38159/ehass.202451119

Aspin, S. H., Ali, M., Abd, M., & Bunyamin, H. (2022). STEM Education in Malaysia: A Review. *Learning Science And Mathematics Journal,* 1(17), 125–139. DOI: https://doi.org/2637-0832

Bahari, F. A. Z., & Saleh, S. (2023). Content Validation Procedure: Development of Problem-solving Skills Test (PSST). *Jurnal Pendidikan Sains Dan Matemaitk Malaysia*, *13*(1), 1–9. DOI: https://doi.org/10.37134/jpsmm.vol13.1.1.2023

Bishop, J. L., & Verleger, M. A. (2013). The Flipped Classroom: A Survey of the Research. *120th American Society for Engineering Education Annual Conference and Exposition*, 30, 1-18.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*(149). DOI: https://doi.org/10.3389/fpubh.2018.00149

Bogaard, G., Dimopoulos, R., & Quaedflieg, C. W. E. M. (2025). Lying Under Pressure: Examining the Impact of Stress and Veracity on Verbal and Nonverbal Behaviors in Credibility Judgments. *Applied Cognitive Psychology*, 39(1). DOI: https://doi.org/10.1002/acp.70018

Cameron, B. T., & Hynes, C. (2024). Expert Panels in Evaluation: An Update From the Field Using the DATA Model. *Canadian Journal of Program Evaluation,* 39(1), 117–143. DOI: https://doi.org/10.3138/cjpe-2023-0047

Cheah, S. M., & Sale, D. (2019). Enhancing Student Engagement In Flipped Classroom Using Autonomy-Supportive Teaching. *Proceedings of the International CDIO Conference.*

Cho, H. J., Zhao, K., Lee, C. R., Runshe, D., & Krousgrill, C. (2021). Active learning through flipped classroom in mechanical engineering: improving students' perception of learning and performance. *International Journal of STEM Education,* 8(1). DOI: https://doi.org/10.1186/s40594-021-00302-2

Chung, E. J., & Lee, B.-H. (2018). The effects of flipped learning on learning motivation and attitudes in a class of college physical therapy students. *Journal of Problem-Based Learning,* 5(1), 29-36. DOI: https://doi.org/10.24313/jpbl.2018.5.1.29

Creswell, J. W. ., & Creswell, J. David. (2022). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (6th ed.). SAGE Publications, Inc.

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

Devellis, R. F. (2016). *Scale Development : Theory and Applications* (4th ed.). SAGE Publications.

Divjak, B., Rienties, B., Iniesto, F., Vondra, P., & Žižak, M. (2022). Flipped classrooms in higher education during the COVID-19 pandemic: findings and future research recommendations. *In International Journal of Educational Technology in Higher Education*. 19(1). DOI: https://doi.org/10.1186/s41239-021-00316-4

Ekinci, F., Bektaş, O., Karaca, M., & Yiğit, K. N. (2023). The flipped learning perception scale: A validity and reliability study. *Education and Information Technologies,* 28(11), 14139–14166. DOI: https://doi.org/10.1007/s10639-023-11731-7

Elian, S. A., & Hamaidi, D. A. (2018). The effect of using flipped classroom strategy on the academic achievement of fourth grade students in Jordan. International *Journal of Emerging Technologies in Learning,* 13(2), 110–125. DOI: https://doi.org/10.3991/ijet.v13i02.7816

Environmental Association for Colleges and Universities (EAUC)). (2025). SDG Accord Report 2025. sdgaccord.org

Eren, E. U., & Kenan Dikilitaş. (2025). Teachers' motivational strategies and student motivation across teaching modalities. *Interactive Learning Environments*, 1–14. DOI: https://doi.org/10.1080/10494820.2025.2601300

Fairhurst, N., Koul, R., & Sheffield, R. (2023). Students' perceptions of their STEM learning environment. *Learning Environments Research,* 26(3). DOI: https://doi.org/10.1007/s10984-023-09463-z

Feijóo, J. C. M., Suárez, F., Chiyón, I., & Alberti, M. G. (2021). Some web-based experiences from flipped classroom techniques in aec modules during the covid-19 lockdown. *Education Sciences,* 11(5). DOI: https://doi.org/10.3390/educsci11050211

George, D., & Mallery, P. (2019). Reliability Analysis. IBM SPSS Statistics 25 Step by Step: A Simple Guide And Reference. In IBM SPSS Statistics 25 Step by Step.

Gong, J., Cai, S., & Cheng, M. (2024). Exploring the Effectiveness of Flipped Classroom on STEM Student Achievement: A Meta-analysis. *Technology, Knowledge and Learning*, 29(2). DOI: https://doi.org/10.1007/s10758-023-09700-7

Hakim Mahari, & Qistina Sallehuddin. (2025). *Nearly half of Form Four students taking STEM subjects*. NST Online; New Straits Times. https://www.nst.com.my/news/nation/2025/10/1300142/nearly-half-form-four-students-taking-stem-subjects retrieved on 23 June 2024

Hertzog, M. A. (2008). Considerations in Determining Sample Size for Pilot Studies. *Research in Nursing & Health*, *31*(2), 180–191. DOI: https://doi.org/10.1002/nur.20247

Intan, S. N. A., & Yasin, S. M. (2021). Content Validity of a Self-Reported Instrument for Safety and Health (S&H) Culture Practice In Paramedic Training Institute Using A Heterogeneous Expert Panel. T*urkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(7), 2464–2472. https://turcomat.org/index.php/turkbilmat/article/view/3574/3052

Ismail, M. H. Bin, Salleh, M. F. M., & Nasir, N. A. M. (2019). The Issues and Challenges in Empowering STEM on Science Teachers in Malaysian Secondary Schools. *International Journal of Academic Research in Business and Social Sciences*, 9(13), 430-444. DOI: https://doi.org/10.6007/IJARBSS/v9-i13/6869

Jamaluddin, F., Razak, A. Z. A., & Rahim, S. S. A. (2025). Navigating the challenges and future pathways of STEM education in Asia-Pacific region: A comprehensive scoping review. STEM Education, 5(1), 53–88. DOI: https://doi.org/10.3934/steme.2025004

Johanson, G. A., & Brooks, G. P. (2010). Initial Scale Development: Sample Size for Pilot Studies. *Educational and Psychological Measurement*, *70*(3), 394–400. DOI: https://doi.org/10.1177/0013164409355692

Kalithas, M., & Manivannan, M. L. (2024). UKM TESL Students' Perceptions and Satisfaction on the Use of Flipped Classroom Approach. *LSP International Journal,* 11(2), 115–125. DOI: https://doi.org/10.11113/lspi.v11.22901

Karadag, R., & Keskin, S. S. (2017). The effects of flipped learning approach on the academic achievement and attitudes of the students. *New Trends and Issues Proceedings on Humanities and Social Sciences,* 4(6). DOI: https://doi.org/10.18844/prosoc.v4i6.2926

Karjanto, N., & Acelajado, M. J. (2022). Sustainable Learning, Cognitive Gains, and Improved Attitudes in College Algebra Flipped Classrooms. *Sustainability* (Switzerland), 14(19). DOI: https://doi.org/10.3390/su141912500

Karnia, R. (2024). Importance of Reliability and Validity in Research. *Psychology and Behavioral Sciences*, *13*(6), 137–141. https://doi.org/10.11648/j.pbs.20241306.11

Kay, R., & Dermott, K. (2018). Examining Stem-Based Flipped Classrooms in Higher Education: A Review of the Literature. *ICERI2018 Proceedings,* 1. DOI: https://doi.org/10.21125/iceri.2018.0076

Kozikoğlu, İ., Erbenzer, E., & Ateş, G. (2021). The Relationship Between Teachers' Perceptions of Flipped Learning Self-efficacy and Their Behaviors to Support Learner Autonomy. *Atatürk Üniversitesi Kazım Karabekir Eğitim Fakültesi Dergisi,* 42, 344–366. DOI: https://doi.org/10.33418/ataunikkefd.796531

Lam, C. P., & Siew, N. M. (2024). Flipped Classroom in Science Education: Correlating Student Experience with Attitudes. Problems of Education in the 21st Century, 82(5), 672–686. https://doi.org/10.33225/pec/24.82.672 Lamm, K., Lamm, A., & Edgar, D. (2020). Scale Development and Validation: Methodology and Recommendations. *Journal of International Agricultural and Extension Education*, *27*(2), 24–35. DOI: https://doi.org/10.5191/jiaee.2020.27224

Lo, C. K., & Hew, K. F. (2019). The impact of flipped classrooms on student achievement in engineering education: A meta-analysis of 10 years of research. Journal of Engineering Education, 108(4), 523-546. DOI: https://doi.org/10.1002/jee.20293

Lukmanul Hakim, N. A. M., Rasidi Pairan, M., & Ikram Zakaria, M. (2025). Step-By-Step Guide to Calculating Content Validity Index (CVI) For Single Constructs Using Excel. *International Journal of Research and Innovation in Social Science*, *IX*(III), 1717–1726. https://doi.org/10.47772/ijriss.2025.90300135

Lynn, M. R. (1986). Determination and quantification of content validity. Nursing Research, 35(6). DOI: https://doi.org/10.1097/00006199-198611000-00017

Ma, Y. (2023). Exploration of flipped classroom approach to enhance critical thinking skills. *Heliyon*, *9*(11), e20895. DOI: https://doi.org/10.1016/j.heliyon.2023.e20895

Martínez-Jiménez, R., & Ruiz-Jiménez, M. C. (2020). Improving students' satisfaction and learning performance using flipped classroom. *International Journal of Management Education,* 18(3). DOI: https://doi.org/10.1016/j.ijme.2020.100422

Mat Said, N. A., Bujang, S. M., Buang, N. A., Siraj@Ramli, H. H., & Awang Besar, M. N. (2022). Critical Thinking Transfer Practice Instrument: A Content Validity Calculation Steps Based on Expert Panel Evaluation. *Education in Medicine Journal,* 14, 61–74. DOI: https://doi.org/10.21315/eimj2022.14.3.5

Md Zainuddin, M. Z. A., Mohd Noor, N. H., Johari, N. 'Aisyah, Othman, E. S., & Mat Fadzil, A. F. (2024). Flipped Classroom and its Effects on Student's Academic Performance and Behaviour. *International Journal of Academic Research in Business and Social Sciences, 1*4(9), 399-415. DOI: https://doi.org/10.6007/IJARBSS/v14-i9/22372

Milligan, M. (2022, September 1). Why are the UN Sustainable Development Goals Important to STEM? Medium. https://abetaccredit.medium.com/why-are-the-un-sustainable-development-goals-important-to-stem-e08a6addfceb

Nja, C. O., Orim, R. E., Neji, H. A., Ukwetang, J. O., Uwe, U. E., & Ideba, M. A. (2022). Students' attitude and academic achievement in a flipped classroom. *Heliyon,* 8(1). DOI: https://doi.org/10.1016/j.heliyon.2022.e08792

Nolan, E., Brady, M., Rienties, B., & Héliot, Y. (2021). Once more on the rollercoaster: Loses and gains from the rapid shift to online delivery during Covid. *Academy of Management Proceeding*s, 2021(1). DOI: https://doi.org/10.5465/ambpp.2021.15358abstract

Nunnally, J., & Bernstein, I. (1994). *Psychometric Theory,* 3rd edn, 1994. McGraw-Hill, New York, 3.

Öztürk, M., & Çakıroğlu, Ü. (2021). Flipped learning design in EFL classrooms: implementing self-regulated learning strategies to develop language skills. *Smart Learning Environments*, 8(1). DOI: https://doi.org/10.1186/s40561-021-00146-x

Pallant, Julie. (2020). *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS* (7th ed.). Routledge: Allen & Unwin. https://www.taylorfrancis.com/books/mono/10.4324/9781003117452/spss-survival-manual-julie-pallant

Pilu, R., Jabu, B., & Sulaiman, I. (2025). Flipped learning and its challenges: understanding students' struggles in Indonesian EFL contexts from teachers' viewpoint. *Frontiers in Education,* 10,. DOI: https://doi.org/10.3389/feduc.2025.1575385

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 and Health*, 29(5), 489-497. DOI: https://doi.org/10.1002/nur.20147

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

Rahman, L. (2024). Pedagogical Approach in STEM Education: A Literature Review. *International Journal of Engineering Research & Technology (IJERT)*, *13*(9). DOI: https://doi.org/10.5281/zenodo.18137298

Rodrigues, I. B., Adachi, J. D., Beattie, K. A., & MacDermid, J. C. (2017). Development and validation of a new tool to measure the facilitators, barriers and preferences to exercise in people with osteoporosis. *BMC Musculoskeletal Disorders,* 18(1), 540. DOI: https://doi.org/10.1186/s12891-017-1914-5

Rubio, D. M., Berg-weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying Content Validity: In Social Work Research. Social Work Research, 27(2).

Sağlam, D., & Arslan, A. (2018). The Effect of Flipped Classroom on the Academic Achievement and Attitude of Higher Education Students. *World Journal of Education*, 8(4), 170-176. DOI: https://doi.org/10.5430/wje.v8n4p170

Segal, D. L., & Coolidge, F. L. (2018). Reliability. In M. H. Bornstein (Ed.), *The SAGE Encyclopedia of Lifespan Human Development.* SAGE Publications, Inc. DOI: https://doi.org/10.4135/9781506307633.n683

Shi, J., Mo, X., & Sun, Z. (2012). Content validity index in scale development. *Journal of Central South University (Medical Sciences),* 37(2), 152-155.. DOI: https://doi.org/10.3969/j.issn.1672-7347.2012.02.007

Shrotryia, V. K., & Dhanda, U. (2019). Content Validity of Assessment Instrument for Employee Engagement. *SAGE Open*, *9*(1). 215824401882175. DOI: https://doi.org/10.1177/2158244018821751

Stefana, A., Damiani, S., Umberto Granziol, Umberto Provenzani, Solmi, M., Youngstrom, E. A., & Paolo Fusar-Poli. (2025). Psychological, psychiatric, and behavioral sciences measurement scales: best practice guidelines for their development and validation. *Frontiers in Psychology*, *15*. DOI: https://doi.org/10.3389/fpsyg.2024.1494261

Sukserm, P. (2024). Determining the Appropriate Sample Size in EFL Pilot Studies. *Journal of Research Methodology*, *37*(3), 245-264. https://doi.org/10.14456/jrm.2024.13

Syaiful, I. A., & Roebianto, A. (2020). Adapting and Examining the Factor Structure of the Self-Compassion Scale in Indonesian Version. *Jurnal Psikologi,* 47(3), 175-205. DOI: https://doi.org/10.22146/jpsi.57608

Syakdiyah, H., Wibawa, B., & Muchtar, H. (2018). The effectiveness of flipped classroom in high school Chemistry Education. IOP Conference Series: Materials Science and Engineering, 434(1). DOI: https://doi.org/10.1088/1757-899X/434/1/012098

Talan, T., & Gulsecen, S. (2019). The effect of a flipped classroom on students' achievements, academic engagement and satisfaction levels. *Turkish Online Journal of Distance Education*, 20(4), 31 - 60. DOI: https://doi.org/10.17718/TOJDE.640503

Unal, A., & Unal, Z. (2021). Using Flipped Classroom in Middle Schools: Teachers' Perceptions. *Journal of Research in Education,* 30(2), 90-112.

UNESCO. (2017). Education for Sustainable Development Goals: learning objectives. UNESCO. DOI: https://doi.org/10.54675/CGBA9153

Warmoth, K., O'Brien, L., Bilciu, A.-M., Dearling, J., Lynch, J., Woolham, J., Mioshi, E., & Almack, K. (2025). The role and contribution of experts by experience in building research capacity in adult social care services: findings from the script study. *Research Involvement and Engagement,* 11(1), 103. DOI: https://doi.org/10.1186/s40900-025-00779-z

Webb, M., & Doman, E. (2020). Impacts of flipped classrooms on learner attitudes towards technology-enhanced language learning. *Computer Assisted Language Learning,* 33(3), 240-274.. DOI: https://doi.org/10.1080/09588221.2018.1557692

Xiaoying, L., & Abu Samah, N. (2024). The Effectiveness, Benefits and Challenges of the Implementation of Flipped Classroom in English Teaching and Learning: A Systematic Review. Sains Humanika, 16(3), 113–121. DOI: https://doi.org/10.11113/sh.v16n3.2164

Yoong, S. M., Hosshan, H., Arumugam, S., Alya Qasdina Ng Ai Lee, Lau, S. C., & Govindasamy, P. (2022). Validity and Reliability of Needs Analysis Questionnaire for Dyscalculia Instrument. *South Asian Journal of Social Science and Humanities*, 3(3), 111-124. DOI: https://doi.org/10.48165/sajssh.2022.3307

Yusoff, M. S. B. (2019). ABC of Content Validation and Content Validity Index Calculation. *Education in Medicine Journal,* 11(2), 49-54. DOI: https://doi.org/10.21315/eimj2019.11.2.6