

# Validity and Reliability of Pre-service Teachers and Teachers' Self-Efficacy: A Scoping Review

Riyan Hidayat<sup>1\*,2</sup>, Ahmad Fauzi Mohd Ayub<sup>1,2</sup>, Harris Shah Abd Hamid<sup>3</sup>, Nurihan Nasir<sup>4</sup>

<sup>1</sup>Faculty of Educational Studies, Universiti Putra Malaysia, 43400 UPM Serdang, Malaysia

<sup>2</sup>Institute for Mathematical Research, Universiti Putra Malaysia, Selangor, Malaysia.

riyan@upm.edu.my

afmy@upm.edu.my

<sup>3</sup>University College MAIWP International, 68100, Batu Caves, Kuala Lumpur, Malaysia

drharris@ucmi.edu.my

<sup>4</sup>Universiti Pendidikan Sultan Idris, 35900, Tanjung Malim, Malaysia

nurihan@fsmt.upsi.edu.my

\*Corresponding Author

<https://doi.org/10.24191/ajue.v21i1.5678>

*Received: 17 September 2024*

*Accepted: 28 December 2024*

*Date Published Online: 11 March 2025*

*Published: 11 March 2025*

**Abstract:** Self-efficacy is defined as an individual's belief in their capacity to organize and carry out the actions required to attain objectives. Academic research has consistently highlighted the increasing significance of self-efficacy for both pre-service and in-service teachers. This scoping review provides a detailed overview of the evaluation of self-efficacy among pre-service and in-service teachers, focusing on the validity measurement approaches, reliability assessment methods, and specific contextual applications. This research reviewed papers in the Scopus, Education Resources Information Centre (ERIC), and ScienceDirect databases regardless of years. From an initial pool of 22,053 documents, 40 studies met the eligibility criteria and were included in the review. The findings show that content and construct validity were frequently emphasized in the research methodologies. Internal consistency coefficients were the most used method for evaluating the reliability of measurement scales. Most studies concentrated on the broader educational field. This review highlights the necessity of a nuanced and comprehensive approach to validity assessment in research on teacher self-efficacy.

**Keywords:** Reliability, Self-Efficacy, Scoping Review, Validity

## 1. Introduction

Scholarly works have consistently highlighted and explored the increasing prominence and recognition of self-efficacy in various fields of study. This well-documented trend underscores its growing importance and relevance in contemporary research and practice. Self-efficacy is an individual's conviction to act in each circumstance to accomplish objectives (Bandura, 1997). Self-efficacy has garnered considerable attention in organizational psychology and educational research (Boulden et al., 2021). A substantial corpus of research has firmly established that teacher self-efficacy exerts a direct impact on multiple facets of education, including student achievement (Perera & John, 2020), the quality of instruction (Zakariya, 2020), student motivation, performance in areas such as mathematics and social-emotional development (Höltge et al., 2019), student engagement levels (Perera & John, 2020) and the quality of

instruction (Burić & Kim, 2020). Regrettably, gender differences had no impact on self-efficacy (Hashim et al., 2022). Therefore, our perspective highlights the central significance of self-efficacy in the realm of education. Notably, there has been a marked surge in research interest concerning the assessment of pre-service teachers and teachers' self-efficacy in recent times (Boulden et al., 2021; Saad et al., 2023).

Task-specific assessments of self-efficacy have proven to be more accurate predictors of actual performance than general assessments. The concept of self-efficacy permits and encourages the development of task-specific measures, leading to the proliferation of such reviews tailored to various tasks. In the context of teaching, Lazarides and Warner (2020) have identified several examples, including the Teacher Self-Efficacy Scale (TSES) by Bandura, the Teacher Sense of Efficacy Scale by Tschannen-Moren and Hoy, and a measure designed by Gibson and Dembo. In addition to these more general teaching self-efficacy scales, some assessments focus on specific aspects of teaching, such as classroom management (see Slater & Mains, 2020). The specificity of the self-efficacy construct has prompted the development of measures for teaching subjects or topics, broadening the scope of applicability. For instance, there are self-efficacy assessments tailored to teaching engineering to K-12 students (Yoon et al., 2014), teaching STEM subjects (Yang et al., 2021), teaching physical education (Block et al., 2013), teaching nutrition (Brenowitz & Tuttle, 2003) and teaching science (Smolleck et al., 2006). This diversification in self-efficacy measures reflects the growing recognition of the importance of domain-specific self-efficacy assessments in education.

Assessments of self-efficacy have a long and comprehensive history, encompassing both teacher and student efficacy over time (Koe et al., 2021). However, it is important to highlight that none of these prior reviews have specifically focused on conducting a comprehensive examination, such as a scoping review, to investigate the various approaches to assessing validity, the methods used for assessing reliability, and the specific domains or contexts considered when evaluating pre-service teachers and teachers' self-efficacy. These aspects of self-efficacy research have yet to be examined to a significant extent. This emphasizes a noticeable gap in the existing body of literature because a thorough exploration of the diverse validity and reliability assessments employed, and the range of domains or contexts taken into consideration when appraising pre-service and teachers' self-efficacy can yield valuable insights and contribute significantly to a more thorough comprehension of this subject matter. Consequently, we propose three research questions below to focus more specifically on teachers' self-efficacy:

1. What types of validity assessments were detailed in the studies?
2. What forms of reliability evaluations were documented in the studies?
3. Which domains or contexts were considered when evaluating pre-service teachers and teachers' self-efficacy?

## **2. Literature Review**

### **2.1 Self-Efficacy**

Self-efficacy, as defined by Bandura in 1977, refers to an individual's confidence in their ability to plan and execute the necessary actions to achieve specific goals. Self-efficacy beliefs, rooted in social cognitive theory, hold significant importance as predictors of behaviour and as a wellspring of motivation (Henson, 2001). These beliefs not only directly influence behaviour but also have broader ramifications. They are positively associated with personal accomplishments, job satisfaction, and commitment while exhibiting a negative relationship with burnout (Zee & Koomen, 2016). In the classroom context, the impact of self-efficacy extends beyond the individuals who hold these beliefs. Teachers' self-efficacy has been linked to students' academic adjustment, achievement, and motivation (Lazarides & Warner, 2021). This highlights the reciprocal influence of personal cognition on environmental factors, in line with Bandura's theoretical framework. Essentially, the confidence teachers have in their abilities plays a pivotal role in their own well-being and in shaping their students' educational experiences and outcomes.

Self-efficacy refers to an individual's intrinsic conviction about their abilities, oriented towards future endeavours, and grounded in the subjective assessment of their capabilities rather than solely relying on observable skills and competencies (Bandura, 1977). This concept highlights that individuals strongly believe in their ability to accomplish their goals, primarily influenced by their internal evaluations of their potential. These assessments surpass the limited showcase of skills to encompass a proactive viewpoint, wherein individuals use their perceived capabilities as a guiding force to shape their future actions and aspirations. Fundamentally, self-efficacy plays a crucial role as a psychological construct impacting an individual's motivation, decision-making, and resilience in the face of obstacles, given its foundations in their self-confidence and belief in their ability to succeed in tasks and attain objectives. Resnick (2008) puts forward the idea that the development of a self-efficacy measurement instrument should be intricately attuned to a particular situation. This approach ensures that the assessment tool effectively gauges an individual's self-assured capacity to perform a specific behaviour and achieve desired outcomes within a clearly defined and relevant context. By customizing self-efficacy assessments, researchers and practitioners can obtain a more accurate and nuanced understanding of an individual's belief in their capabilities and how these beliefs manifest in real-world scenarios.

## 2.2 Validity and Reliability in Research

Validity encompasses several dimensions, including appropriateness, meaningfulness, accuracy, and utility. These factors together determine the overall quality and reliability of the conclusions drawn by a researcher (Hidayat, 2024). This study used standard definitions for different validity and reliability measures as the conceptual framework to extract and interpret data (Table 1) (Ratanawongsa et al., 2008). These established definitions served as the foundation for our analysis and ensured that we followed widely accepted guidelines for assessing the trustworthiness and accuracy of the data we gathered.

**Table 1**

*The Definitions for Reliability and Validity Applied in the Scoping Review*

Category	Psychometric property	Definitions
Face validity		The extent to which the items and questions presented are clear, allow sufficient time for responses, and, most crucially, assess what they are intended to assess
Content validity		The extent to which the questions in an instrument and the scores derived from those questions represent the content
Construct validity		The extent to which an assessment accurately assesses the theoretical concept it is designed to measure
Discriminant validity		The extent to which measurements of constructs that have no theoretical connection are not influenced by each other

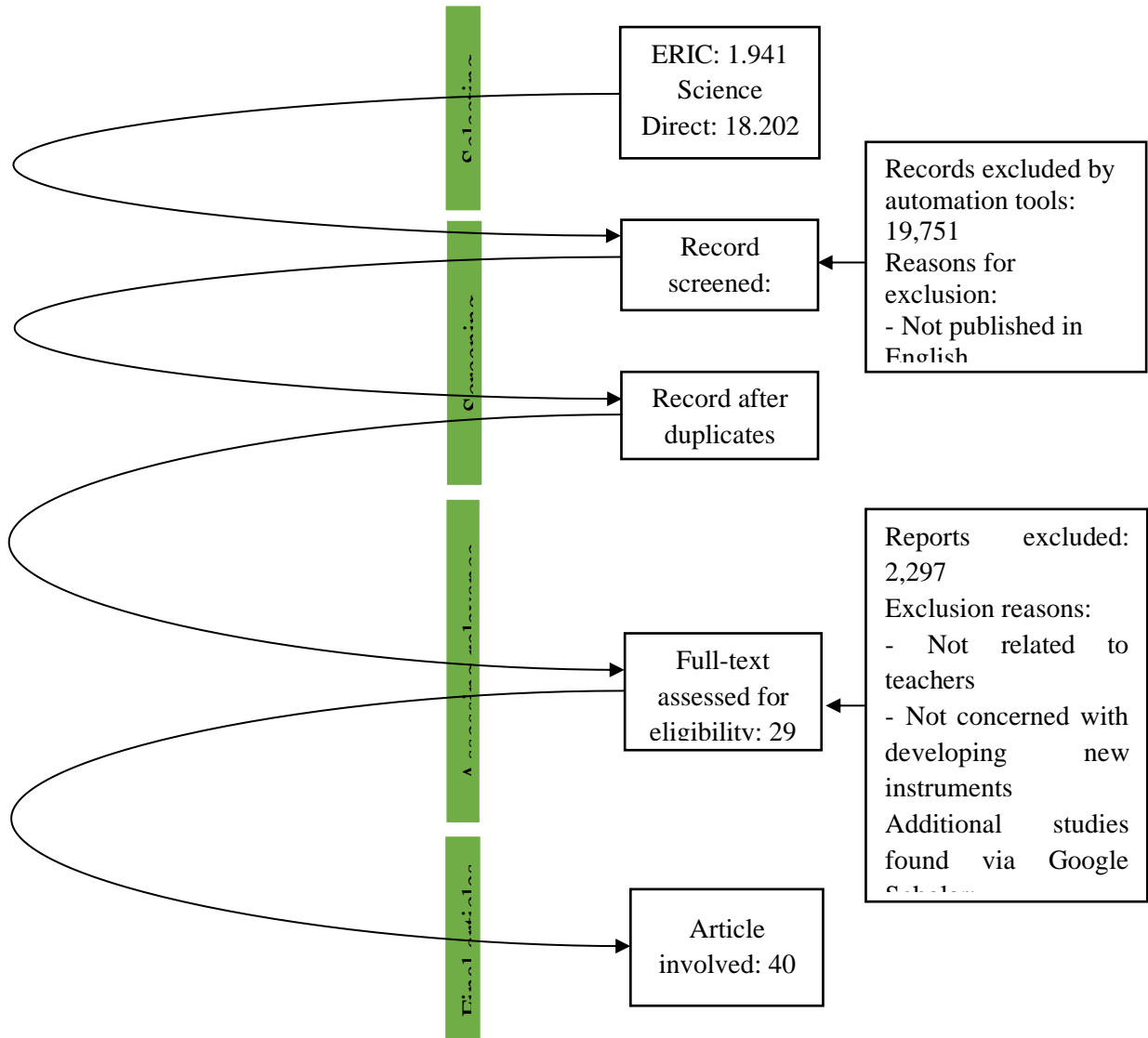
Category	Psychometric property	Definitions
Convergent validity		An approximation of the association between measurements of constructs that have a theoretical connection
Criterion validity	Concurrent validity	The extent to which a tool generates results that align with those of a recognized or validated instrument measuring the same variable
	Predictive validity	The extent to which a measurement accurately foretells anticipated results
Reliability	Inter-rater reliability	The extent to which measurements remain consistent when collected by different individuals
	Intra-rater reliability	The extent to which measurements exhibit consistency when repeatedly taken by the same individual
	Test–retest reliability	The extent to which consistent results are obtained when the same test is repeated under identical conditions
	Internal consistency	The extent to which items representing the same concept produce comparable outcomes

### 3. Method

The methodology of a scoping review involves a structured approach comprising five key stages: 1) formulating the research question, 2) identifying relevant studies, 3) selecting those studies, 4) organizing and documenting the data, and 5) aggregating, summarizing, and presenting the findings (Utami et al., 2024). Another distinguishing feature of a scoping review is its ability to identify different types of existing evidence within a particular field and clarify crucial concepts or definitions found in the literature (Munn et al., 2018). It offers an ideal framework to explore the conceptualization of the pre-service teachers and teachers' self-efficacy construct, assess the utilization of self-efficacy theories, and comprehensively examine the validity of assessment instruments. We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) protocol (Tricco et al., 2018). We adhered to its flowchart (Figure 1).

**Figure 1**

Selection Process



3.1 Selecting

We employed three prominent search engines, namely Education Resources Information Centre (ERIC), ScienceDirect, and Scopus. We opted for these repositories due to their stringent indexing standards, esteemed global reputation, and comprehensive coverage of teacher education research. We manually examined the reference lists of identified articles and utilized Google Scholar to explore additional primary sources in the grey literature. We also applied backward and forward tracking techniques. In the case of backward tracking, we manually retrieved the reference lists of the initial references used as seeds. For forward tracking, we employed Google Scholar to identify citing references, which are sources that reference the initial seed references. The subsequent step involved the identification of keywords to locate relevant journal articles using three distinct search engines. We derived suitable keywords from the article titles to facilitate our review, which included terms such as self-efficacy, teacher, educator, instrument, and measurement. These keywords were combined using Boolean operators like "OR" and "AND" to create a search string, serving as the basis for our literature search strategy during the identification phase. We searched the ScienceDirect and Scopus databases using the keywords

("self-efficacy") AND ("instrument" OR "measurement" OR "tool") AND ("teacher" OR "educator"). For the ERIC database, we applied the keywords "self-efficacy\*" AND "instrument" OR "measurement" OR "tool" AND "teacher" OR "educator". This process yielded a total of 22,053 articles across the ERIC, ScienceDirect, and Scopus databases.

### **3.2 Screening**

Before removing duplicate papers, we conducted an automated screening process. In assessing the titles, our specific criterion was that they should relate to self-efficacy within the context of teacher education. Given that our chosen topic revolved around pre-service teachers and teachers' self-efficacy, we refined our focus to encompass only in-service teachers as the primary subjects of the articles we sought while excluding pre-service teachers. Additionally, this screening process entailed the exclusion of papers not published in the English language and those that had yet to undergo the peer-review process, ensuring that the final selection comprised high-quality, peer-reviewed articles in English. During this phase, we systematically excluded 19,751 articles that did not meet our criteria of being written in English and not having undergone peer review. Among the initial 19,751 results, we identified and subsequently removed five duplicate findings from the dataset. This curation process aided in streamlining the dataset for further analysis. We proceeded to the next stage after removing these five duplicate records using Excel software.

### **3.3 Evaluating Relevance**

We meticulously evaluated the titles, abstracts, and keywords of 2,297 papers following our predefined inclusion and exclusion criteria. This rigorous assessment led us to identify 30 articles aligned with our eligibility criteria. During the relevance assessment phase, we considered articles offering full-text access. In this review, our initial criteria primarily emphasized the inclusion of articles exclusively from peer-reviewed journals. Our second criterion involved the inclusion of studies that met two specific conditions. Firstly, they needed to assess self-efficacy using newly developed assessment tools. Secondly, the participants in these studies had to be pre-service and teachers. Additionally, the third requirement was to incorporate papers published in English. Prioritizing English journal publications was intended to mitigate the potential challenges associated with complex or ambiguous translations. As a result, after removing articles that did not meet the criteria for inclusion in this scoping review, we retained 30 full-text articles while excluding 2,297 papers from consideration.

### **3.4 Final Sample**

While examining the full texts of the studies, we employed backward and forward tracking to identify additional relevant research. Backward tracking involved tracing the references cited in the selected studies (five) to uncover related research, while forward tracking entailed exploring newer studies that had referenced the selected ones (six). This approach allowed us to discover valuable research that might have been overlooked and augmented the pool of relevant studies for our analysis. We deemed these 11 newly identified studies pertinent and subsequently integrated them into the existing 40 identified studies. This combined collection of 40 studies is the foundation for our research analysis, encompassing a more extensive range of literature for our review and examination. The authors can confirm that 40 studies have met the criteria.

### **3.5 Data Analysis**

We conducted a thematic analysis of the qualitative data using an inductive approach, enabling us to identify and emphasize prominent themes (Nowell et al., 2017). To streamline data management, we created a data extraction table and compiled all relevant information from the included studies within a Microsoft Excel spreadsheet. To ensure the validity of the research, a collaborative effort involving all four authors was employed to develop and categorize themes based on shared characteristics and relevance. When coding complexities arose, the authors engaged in discussions and, if necessary, sought independent coding from other authors to resolve any discrepancies or uncertainties. After a comprehensive review of the 40 retained papers, all authors reached a consensus.

#### 4. Results

Table 2 provides an overview of the articles that satisfied our selection standards and, as a result, were integrated into this review.

**Table 2**

*Research Articles Incorporated in the Scoping Review*

Author(s)	Year	Types of sample	Context	Type of validity	Type of reliability
Guskey	1981	In-service teachers	General domain	✓ Construct validity using factor analysis	✓ Internal consistency (Cronbach's alpha, 0.79) ✓ Unequal-length Spearman Brown (0.76) ✓ Guttman split half (0.75)
Rose & Medway	1981	In-service teachers	General domain	✓ Construct validity using principal-factoring solution	✓ Internal consistency (Kuder Richardson, 0.71 - 0.81)
Betz & Hackett	1983	Pre-service teachers	Mathematics domain	✓ Not clearly reported	✓ Internal consistency (Cronbach's alpha, 0.96)
Gibson & Dembo	1984	In-service teachers	General domain	✓ Content validity by experts and literature ✓ Construct	✓ Internal consistency (Cronbach's alpha, 0.75 - 0.79)

Author(s)	Year	Types of sample	Context	Type of validity	Type of reliability
				validity using factor analysis ✓ Convergent validity using bivariate correlations ✓ Discriminant validity using bivariate correlations	
Riggs & Enochs	1990	Pre-service teachers	Science domain	✓ Content validity by experts ✓ Construct validity using confirmatory factor analysis	✓ Internal consistency (Cronbach's alpha, 0.76 - 0.90)
Bandura	1997	In-service teachers	General domain	✓ Content validity generated based on literature ✓ Construct validity (not clearly reported)	✓ Internal consistency (Cronbach's alpha, not clearly reported)
Kranzler & Pajares	1997	Pre-service teachers	Mathematics domain	✓ Construct validity using principal components analysis	✓ Internal consistency (Cronbach's alpha, 0.91 - 0.95)
Enochs, Smith & Huinker	2000	Pre-service teachers	Mathematics domain	✓ Construct validity using confirmatory factor analysis	✓ Internal consistency (Cronbach's alpha, 0.77 - 0.88)
Roberts & Henson	2000	In-service teachers	General domain	✓ Construct validity using confirmatory factor analysis	✓ Not clearly reported
Tschannen-Moran & Hoy	2001	In-service	General domain	✓ Construct validity using	✓ Internal consistency



Author(s)	Year	Types of sample	Context	Type of validity	Type of reliability
		teachers		principal-axis factoring ✓ Discriminant validity using bivariate correlations	(Cronbach's alpha, 0.87 - 0.91)
Martin & Kulinna	2003	In-service teachers	Physical domain	✓ Construct validity using exploratory and confirmatory factor analysis	✓ Internal consistency (Cronbach's alpha, 0.73 - 0.86)
Heneman III, Kimball & Milanowski	2006	In-service teachers	General domain	✓ Construct validity using confirmatory factor analysis	✓ Internal consistency (Cronbach's alpha, 0.91) ✓ Composite reliability (0.72)
Skaalvik & Skaalvik	2007	In-service teachers	General domain	✓ Construct validity using confirmatory factor analysis	✓ Internal consistency (Cronbach's alpha, 0.61 - 0.89)
Dellinger, Bobbett, Olivier & Ellett	2008	In-service teachers	Mathematics domain	✓ Principal component analysis (not clearly reported)	✓ Reliability analysis (not clearly reported)
Teo	2009	Pre-service teachers	General domain	✓ Construct validity using confirmatory factor analysis ✓ Convergent validity using average variance extracted (AVE) ✓ Discriminant validity using the square root	✓ Composite reliability (0.86 - 0.89)

Author(s)	Year	Types of sample	Context	Type of validity	Type of reliability
				of AVE and correlations	
Tschannen-Moran, & Johnson	2011	In-service teachers	Literacy Domain	<ul style="list-style-type: none"> <li>✓ Content validity by experts</li> <li>✓ Construct validity using exploratory and confirmatory factor analysis</li> <li>✓ Concurrent validity using bivariate correlations</li> </ul>	<ul style="list-style-type: none"> <li>✓ Internal consistency (Cronbach's alpha, 0.96)</li> </ul>
De Paul	2012	In-service teachers	General domain	<ul style="list-style-type: none"> <li>✓ Content validity by experts</li> </ul>	<ul style="list-style-type: none"> <li>✓ Internal consistency (Cronbach's alpha, 0.94)</li> <li>✓ Split Half method (0.90)</li> </ul>
McGee & Wang	2014	In-service teachers	Mathematics domain	<ul style="list-style-type: none"> <li>✓ Content validity by experts</li> <li>✓ Construct validity using exploratory and confirmatory factor analysis</li> </ul>	<ul style="list-style-type: none"> <li>✓ Internal consistency (Cronbach's alpha, 0.86 - 0.93)</li> </ul>
Yoon, Evans & Strobel	2014	In-service teachers	Engineering domain	<ul style="list-style-type: none"> <li>✓ Face validity by experts</li> <li>✓ Content validity by experts</li> <li>✓ Construct validity using exploratory and confirmatory factor analysis</li> </ul>	<ul style="list-style-type: none"> <li>✓ Internal consistency (Cronbach's alpha, 0.89 - 0.96)</li> </ul>

Author(s)	Year	Types of sample	Context	Type of validity	Type of reliability
				✓ Discriminant validity using confirmatory factor analysis	
Buns & Thomas	2015	In-service teachers	Physical education domain	<ul style="list-style-type: none"> <li>✓ Concurrent validity using Spearman's correlation</li> <li>✓ Discriminant validity (Pearson's product-moment correlation)</li> </ul>	<ul style="list-style-type: none"> <li>✓ Internal consistency (Cronbach's alpha, 0.96)</li> <li>✓ Equal-Length Spearman Brown split-half (<math>r = 0.87</math>)</li> <li>✓ Guttman's split-half (<math>r = 0.86</math>)</li> </ul>
Dybowski, Kriston & Harendza	2016	Pre-service teachers	Physicians' clinical domain	<ul style="list-style-type: none"> <li>✓ Construct validity using exploratory structural equation modelling (ESEM) and confirmatory factor analysis</li> <li>✓ Concurrent validity using bivariate correlations</li> <li>✓ Content validity generated based on literature</li> </ul>	<ul style="list-style-type: none"> <li>✓ Internal consistency (Cronbach's alpha, 0.77 - 0.90)</li> </ul>
Park, Dimitrov, Das & Gichuru	2016	Pre-service teachers	Intellectual Disability domain	<ul style="list-style-type: none"> <li>✓ Construct validity using exploratory and confirmatory factor analysis</li> </ul>	<ul style="list-style-type: none"> <li>✓ Internal consistency (Cronbach's alpha, 0.97)</li> </ul>
Dođru	2017	In-service teachers	Technology domain	<ul style="list-style-type: none"> <li>✓ Construct validity using</li> </ul>	<ul style="list-style-type: none"> <li>✓ Internal consistency</li> </ul>

Author(s)	Year	Types of sample	Context	Type of validity	Type of reliability
				exploratory and confirmatory factor analysis	(Cronbach's alpha, 0.93) ✓Test retest method ( $r = 0.83$ )
Veldman, Admiraal, Mainhard, Wubbels & Van Tartwijk	2017	In-service teachers	Interpersonal domain	✓Construct validity using exploratory factor analysis ✓Predictive validity using bivariate correlation ✓Concurrent validity using bivariate correlation	✓Internal consistency (Cronbach's alpha, 0.78 - 0.80)
Demirci & Ozyurek	2018	In-service teachers	Astronomy domain	✓Content validity by experts ✓Construct validity using expert, exploratory and confirmatory factor analysis	✓Internal consistency (Cronbach's alpha, 0.84)
Wilkerson, Eddy, Quebec Fuentes, Sorto, Gupta, Ward & Kerschen	2018	Pre-service teachers	Mathematics domain	✓Content validity by experts ✓Construct validity using exploratory factor analysis	✓Not clearly reported
Handtke & Bögeholz	2019	In-service teachers	Science domain	✓Content validity generated based on literature ✓Construct validity using exploratory and confirmatory	✓Not clearly reported

Author(s)	Year	Types of sample	Context	Type of validity	Type of reliability
				factor analysis ✓ Concurrent validity using bivariate correlation	
Höltge, Ehm, Hartmann & Hasselhorn	2017	In-service teachers	Language, social-emotional, and mathematical domain	✓ Construct validity using exploratory and confirmatory factor analysis ✓ Content validity generated based on literature	✓ Internal consistency (Cronbach's alpha, 0.82 - 0.93)
Handtke & Bögeholz	2020	Pre-service teachers	Science domain	✓ Content validity using experts ✓ Convergent using bivariate correlations ✓ Construct using exploratory and confirmatory factor analysis ✓ Concurrent using bivariate correlations	✓ Internal consistency (Cronbach's alpha, 0.93 - 0.94)
Nazari	2020	Pre-service teachers	Intellectual Disability domain	✓ Content validity using experts ✓ Construct using exploratory and confirmatory factor analysis ✓ Concurrent using bivariate correlations ✓ Convergent validity using average extracted variance (AVE)	✓ Internal consistency (Cronbach's alpha, 0.88) ✓ Composite reliability (CR=0.89)
Alkharusi,	2021	In-	Homework	✓ Content	✓ Internal

Author(s)	Year	Types of sample	Context	Type of validity	Type of reliability
Aldhafri, Al-Harthy, Albarashdi, Alrajhi & Alhadabi		service teachers	Management domain	validity by expert	consistency (Cronbach's alpha, 0.89)
Boulden, Rachmatullah, Oliver & Wiebe	2021	In-service teachers	Computational thinking domain	<ul style="list-style-type: none"> <li>✓ Face validity by expert</li> <li>✓ Content validity by expert</li> <li>✓ Construct validity using Rasch (Differential Item Functioning (DIF) and confirmatory factor analysis</li> </ul>	<ul style="list-style-type: none"> <li>✓ Internal consistency (Cronbach's alpha, 0.78 - 0.95)</li> <li>✓ Person (0.77 - 0.91) and item separation = 0.92)</li> </ul>
Weerasekara, Oh, Cho, & Im	2021	In-service teachers	Nursing domain	<ul style="list-style-type: none"> <li>✓ Content validity using experts</li> <li>✓ Convergent validity using average variance extracted (AVE) and construct reliability (CR)</li> <li>✓ Discriminant validity using average variance extracted (AVE)</li> <li>✓ Construct validity using exploratory factor analysis</li> </ul>	<ul style="list-style-type: none"> <li>✓ Internal consistency (Cronbach's alpha, 0.88 - 0.97)</li> <li>✓ Split-half reliability (0.94 - 0.96)</li> </ul>
Yang, Wu & Li	2021	Pre-service teachers	STEM domain	✓ Content validity generated	✓ Internal consistency (McDonald's

Author(s)	Year	Types of sample	Context	Type of validity	Type of reliability
				based on literature ✓ Construct validity using exploratory factor analysis ✓ Criterion validity using Pearson correlation	s omega ( $\omega$ ), 0.90 – 0.94)
Bal, Yilmaz & Atas	2022	In-service teachers	General domain	✓ Content validity by expert ✓ Construct validity using exploratory and confirmatory factor analysis	✓ Internal consistency (Cronbach's alpha, 0.92) ✓ Guttman Split Half tests (0.78)
Sánchez-Rosas, Dyzenchouz, Dominguez-Lara & Hayes	2022	In-service teachers	General domain	✓ Content validity using experts ✓ Criterion validity using bivariate correlations ✓ Construct validity using exploratory structural equation modeling	✓ Internal consistency (Omega ( $\omega$ ), 0.81) ✓ Internal consistency (Cronbach's alpha, 0.79)
Unfried, Rachmatullah, Alexander & Wiebe	2022	In-service teachers	STEM domain	✓ Content validity using experts ✓ Construct validity using Rasch and confirmatory factor analysis	✓ Internal consistency (Cronbach's alpha, 0.77 - 0.93) ✓ Separation reliability (0.99)
Vatou, Gregoriadis, Tsigilis &	2022	In-service teachers	Social domain	✓ Content validity by experts	✓ Internal consistency (omega

Author(s)	Year	Types of sample	Context	Type of validity	Type of reliability
Grammatikopoulos				✓ Construct validity using exploratory and confirmatory factor analysis ✓ Concurrent validity using intraclass correlation coefficient (ICC)	hierarchical coefficient ( $\omega$ )
Wu, Tseng, Chen, Tseng & Pai	2022	Pre-service teachers	Clinical nursing domain	✓ Content validity by experts ✓ Construct validity using exploratory factor analysis	✓ Internal consistency (Cronbach's alpha, 0.83 - 0.93)
O'Neill	2023	In-service teachers	General domain	✓ Content validity by experts ✓ Construct validity using exploratory factor analysis	✓ Internal consistency (Cronbach's alpha, 0.85 - 0.92)

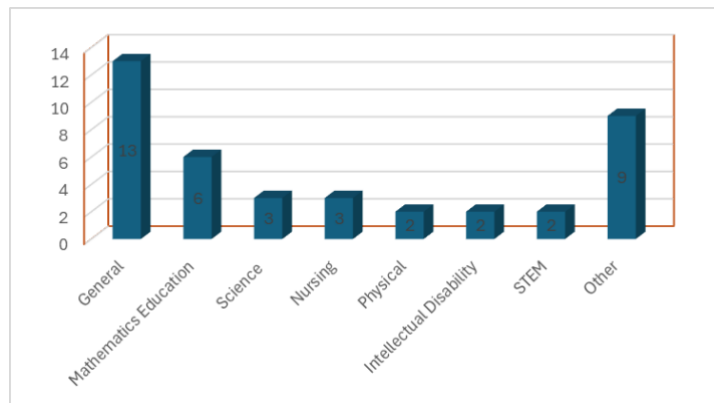
#### 4.1 Context of self-efficacy assessment

Figure 2 offers a synopsis of the principal areas or settings that earlier studies explored in their self-efficacy examinations. Most of these studies, totalling 13 articles, have primarily focused on the general domain, particularly on educational contexts. The general domain refers to teachers' efficacy regarding various aspects of their profession, such as teaching, content delivery, and assessment, without focusing on any specific subject area. To be more precise, six studies have examined mathematics, three have focused on science, and another three on nursing. There are also two studies each on physical disabilities and STEM fields. Additionally, nine articles have explored a range of other topics, including literacy, engineering, clinical practice in medicine, technology, interpersonal relationships, astronomy, language skills, socioemotional and mathematical abilities, homework management, computational thinking, social dynamics, and clinical nursing.

**Figure 2**

*Context of Self-Efficacy Assessment*





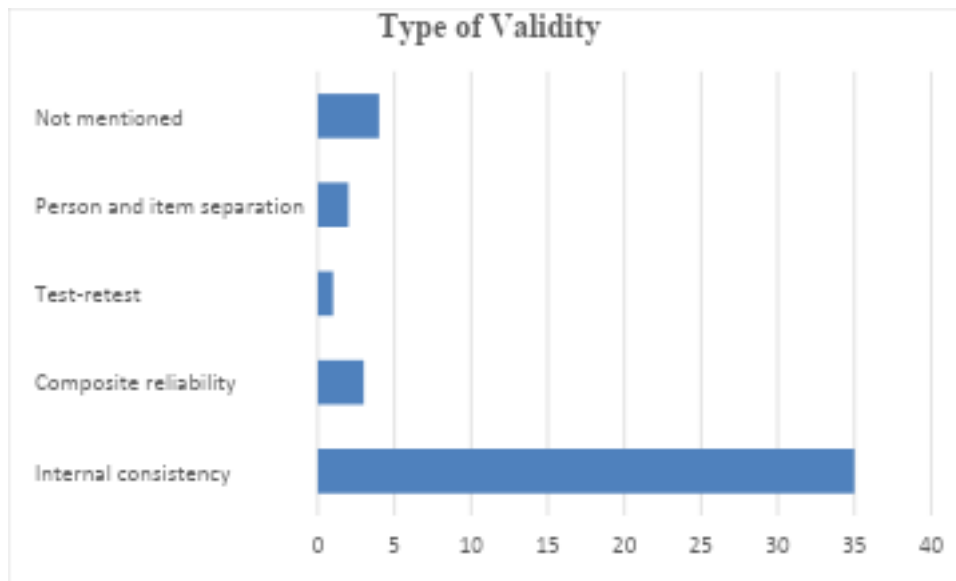
## 4.2 Type of Validity

Figure 3 presents an overview of the types of validity explored in earlier studies on pre-service teachers and teachers' self-efficacy. It identifies face and content validity as key aspects. However, the scoping review found that only two out of 40 articles (5%) discussed face validity, while information on content validity was available in 23 out of 40 articles (57.5%). These articles used various approaches, with 14 focusing on expert judgment and five aligning their measures with existing literature. Only one study combined expert judgment and literature-based criteria for content validity assessment. Construct validity was assessed using diverse methodologies across studies. Twelve out of 40 studies (30%) employed EFA and CFA. Six articles (15%) used either EFA or CFA exclusively. Different techniques, such as factor analysis, principal components analysis, and exploratory structural equation modelling, were also employed. However, one article should have specified its method for construct validity assessment, highlighting the variety of approaches in this area.

Criterion validity was reported in 11 out of 40 studies (27.5%). Among them, eight studies (72.73%) used concurrent validity, while one (9.09%) used predictive validity, relying on bivariate correlation. Two articles (18.18%) provided a more general mention of criterion validity without specifying the method used, demonstrating the diversity in how researchers approach this aspect. Regarding discriminant and convergent validity, only a minority of studies focused on these aspects. Five articles (12.5%) evaluated convergent validity using statistical analyses like bivariate correlations and average variance extracted (AVE). Six articles (15%) assessed discriminant validity through bivariate correlations, AVE, and confirmatory factor analysis (CFA). However, most research (72.5%) did not explicitly address these types of validity, indicating varied research priorities.

**Figure 3**

*Type of Validity*



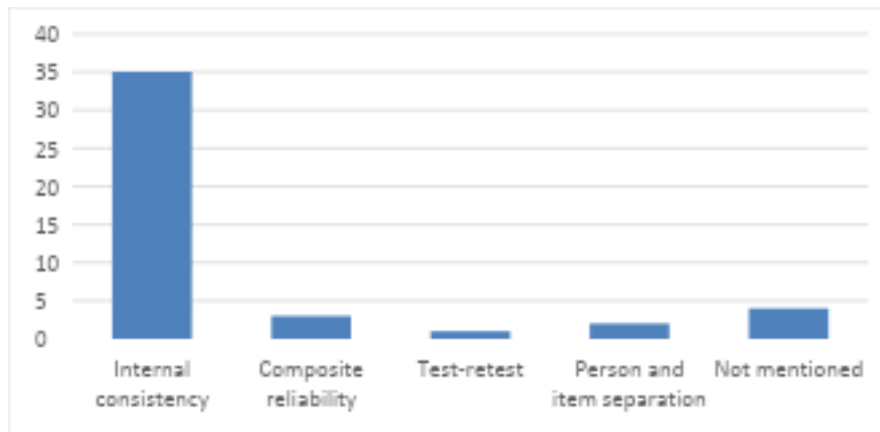
### 4.3 Types of Reliability

In the context of assessing reliability, a significant majority of research, comprising 35 articles or 87.5%, places a strong emphasis on internal consistency (Figure 3). Various statistical analyses are employed to gauge this internal consistency, with methods including Cronbach's Alpha ( $\alpha$ ), Kuder-Richardson Formula 20 (KR-20), Unequal-length Spearman-Brown, Guttman split-half, Split Half method and McDonald's omega ( $\omega$ ) being utilized. These analyses help researchers determine how well the items within a measurement instrument or test correlate and align in measuring the same underlying construct. At the same time, a smaller portion of the research, specifically three articles (7.5%), reports the use of composite reliability. This approach assesses the reliability of a composite or latent variable constructed from multiple indicators.

Additionally, only one article (2.5%) emphasizes test-retest reliability, which examines the stability of measurements over time. Interestingly, two articles (7.5%) adopt the Rasch measurement model to evaluate reliability, specifically focusing on person and item separation. This model provides insights into separating individuals or items along a latent trait continuum. However, it is worth noting that in a subset of four articles (10%), no specific reliability indices or statistical analyses related to reliability are reported in their research findings. This diversity in the approaches and the presence of research that does not explicitly report reliability indices underscore the variability in how researchers address and communicate reliability assessments in their studies.

**Figure 4**

*Type of Reliability Assessment*



## 5. Discussion

Our research findings have shed light on the prevalent practices in the academic landscape. Many research endeavours predominantly incorporate content and construct validity into their methodologies (Siraj et al., 2022; Zhuofan et al., 2024). However, it is noteworthy that a relatively small segment of these studies embraces a more comprehensive approach to validity, encompassing content, construct, face, discriminant, convergent, and criterion (both concurrent and predictive) validity. The limited focus on broader validation measures carries practical implications, underscoring the necessity for more robust assessment tools that effectively assess the strength and relevance of self-efficacy measurements across different educational domains. Closing this gap could increase the reliability of research outcomes and enhance the practical application of these assessments in guiding teacher development and improving instructional practices. When it comes to content validity, scholars exhibit a range of preferences. Some rely exclusively on expert judgment to establish the validity of their content, while others place their trust solely in literature-based validity criteria. Additionally, a subset of researchers adopts a hybrid approach, blending expert judgment and literature-based criteria to ensure robust content validity. In construct validity, two statistical analysis methods, EFA and CFA, emerge as the go-to tools. Researchers frequently employ these techniques individually or in tandem to evaluate the construct validity of their measures. Another alternative method for assessing construct validity involves utilizing the Rasch measurement model. As a result, although content and construct validity are crucial foundations of research, the underutilization of other forms of validity deserves recognition and intervention within the research community. A more inclusive approach to validating research promises to improve the trustworthiness, dependability, and influence of research results in diverse fields.

Criterion validity, conversely, garners more attention in terms of concurrent validity than predictive validity among researchers. Bivariate correlation emerges as the preferred assessment method for concurrent validity in many studies. Moreover, some research studies delve into discriminant and convergent validity, utilizing techniques such as bivariate correlations and the calculation of AVE to ensure the robustness of their measures. However, our scoping review intriguingly revealed that only a few studies incorporate face validity into their research designs. This finding aligns with the perspective of Ratanawongsa et al. (2008), who argue that face validity is no longer regarded as a distinct category. This highlights the evolving nature of validity assessment in contemporary research practices. Therefore, there is an urgent need for more comprehensive approaches to assess validity when developing new measures of pre-service teachers' and teachers' self-efficacy. These methods should be carefully and transparently documented to address this critical issue, with particular emphasis on improving criterion validity.

Our scoping study has revealed a prevalent pattern in assessing the reliability of measurement scales, where the internal consistency coefficient, often referred to as Cronbach's

alpha, stands out as the most frequently used method. In the context of our research, reliability pertains to the extent of steadiness and uniformity observed in the scores generated by the measuring instrument. Importantly, this approach to evaluating reliability has also been widely favoured in reviews conducted across diverse academic domains (Raykov & Grayson, 2003). However, it is noteworthy that a smaller subset of research studies has chosen to employ an alternative approach known as composite reliability. This method emphasizes evaluating the reliability of a composite or latent variable, which is constructed by combining multiple indicators or constituent components. Composite reliability is highly recommended for future use as it provides an index that reflects the impact of measurement error on the scale (Raykov & Grayson, 2003). The distinctive value of composite reliability lies in its ability to offer a more comprehensive viewpoint when assessing the reliability of intricate constructs. It recognizes the interconnectedness and interdependencies among the various measurement items, thereby considering how these items collectively contribute to the overall reliability and stability of the measuring instrument. This nuanced reliability assessment can be especially advantageous when dealing with complex, multifaceted, and multidimensional constructs, enhancing our understanding of the reliability of such measurement tools. Therefore, employing composite reliability enables educators and researchers to ensure that self-efficacy tools are consistently reliable across various contexts and populations, resulting in more dependable outcomes. This, in turn, facilitates more informed decision-making in teacher development programs and instructional strategies, as the data obtained will be more stable and accurately represent true self-efficacy levels.

Most of the studies focused on the broad educational domain or context. In assessing self-efficacy in the field of education, it is a widely recognized concept that unfolds gradually over time. A recent investigation by O'Neill in 2023 underscores the significance of educators possessing self-efficacy who remain well-informed about emerging trends and research. These educators demonstrate a greater willingness to experiment with innovative approaches in their teaching practices, leading to positive outcomes for student achievement, regardless of the educational context, including situations involving military children. However, it is imperative to develop assessment tools specifically tailored to the unique context of pre-service and teacher self-efficacy. This necessity arises because personal attributes such as competence and self-efficacy, environmental influences, and observable behaviours differ among individuals within specific educational domains. To ensure the validity of assessments within these domains, it is crucial to consider factors like fit indices, simplicity in factorial structure, variance explained on average, and internal consistency (Hidayat et al., 2021).

## **6. Conclusion**

Self-efficacy has gained significant attention in organizational psychology and educational research. It is important to highlight that previous reviews did not conduct a comprehensive analysis, such as a scoping review, to investigate the various methods utilized for assessing validity, evaluating reliability, and considering the specific contexts in which pre-service and teachers' self-efficacy is measured. Only a limited number of articles directly addressed the measurement of pre-service teachers and teachers' self-efficacy. Consequently, we had to include earlier studies specifically dealing with this aspect. In our scoping review, many research endeavours primarily incorporate content and construct validity in their methodologies. However, a smaller portion of these studies take a more holistic approach to validity, encompassing content, construct, face, discriminant, convergent, and criterion (both concurrent and predictive) validity. Our scoping study has revealed a common trend in assessing the reliability of measurement scales, with internal consistency coefficients, often referred to as Cronbach's alpha, being the most frequently used method. Lastly, most studies examined the broad educational domain or context. Based on our findings, researchers and educators need to collaborate to enhance and expand the use of validity in research methodologies. Researchers can leverage educators' practical insights and experiences to adapt methodologies to real-world

scenarios, while educators can benefit from researchers' specialized knowledge in advanced validation methods. This joint effort not only improves the development of research tools and techniques but also ensures that findings are accurately applied in practice, leading to more informed and effective educational practices and policies.

## 7. Limitation, Recommendation and Implication

One limitation of this research is its exclusive focus on creating new self-efficacy assessment tools, which, while innovative, overlooks the equally important aspect of using or adapting existing instruments. Many studies customize items from existing tools to suit specific contexts, a practice yet to be explored here. Additionally, while insightful for this group, this study's emphasis on pre-service teachers and educators may only partially apply to other professions or situations. A valuable future research direction is a comparative study evaluating the advantages and disadvantages of developing new self-efficacy assessment tools versus adopting or adapting existing ones, considering factors like cost-effectiveness, reliability, and validity. To broaden our understanding, future research should encompass diverse occupational groups, including pre-service teachers, and explore self-efficacy in professions beyond education, unveiling unique dynamics and commonalities across different fields. The implications drawn from our scoping review regarding the validity evaluation in studies measuring pre-service teachers and teachers' self-efficacy hold significant relevance for future research endeavours. Our review underscores the importance of adopting a nuanced and all-encompassing approach when assessing validity in studies centred on pre-service teachers and teachers' self-efficacy. Subsequent research efforts can capitalize on these findings to elevate the precision and rigour of self-efficacy measurement within educational contexts. Future research should consider a more comprehensive investigation into face validity, recognizing its pivotal role in ensuring that self-efficacy measures resonate as pertinent and suitable for pre-service teachers and educators.

## 8. Funding Acknowledgement

This research received support from the Geran Putra GP-IPM [Grant Numbers: GP-IPM /2023/9744000].

## 9. Co-author Contribution

**Riyan Hidayat:** conceptualization, formal analysis, funding acquisition, methodology, & visualization.

**Riyan Hidayat & Ahmad Fauzi Mohd Ayub:** writing—original draft & writing—review & editing.

**Nurihan Nasir & Harris Shah Abd Hamid:** supervision.

**Ahmad Fauzi Mohd Ayub & Nurihan Nasir:** validation.

**Harris Shah Abd Hamid:** data curation & software.

**Riyan Hidayat & Nurihan Nasir:** investigation & resources.

All authors have agreed with the results and conclusions.

## 10. References

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>

- Block, M. E., Hutzler, Y., Barak, S., & Klavina, A. (2013). Creation and validation of the self-efficacy instrument for physical education teacher education majors toward inclusion. *Adapted Physical Activity Quarterly*, 30(2), 184–205. <https://doi.org/10.1123/apaq.30.2.184>
- Boulden, D. C., Rachmatullah, A., Oliver, K. M., & Wiebe, E. (2021). Measuring in-service teacher self-efficacy for teaching computational thinking: Development and validation of the T-STEM CT. *Education and Information technologies*, 26(4), 4663-4689. <https://doi.org/10.1007/s10639-021-10487-2>
- Brenowitz, N., & Tuttle, C. R. (2003). Development and testing of a nutrition-teaching self-efficacy scale for elementary school teachers. *Journal of Nutrition Education and Behavior*, 35(6), 308–311. [https://doi.org/10.1016/s1499-4046\(06\)60345-x](https://doi.org/10.1016/s1499-4046(06)60345-x)
- Burić, I., & Kim, L. E. (2020). Teacher self-efficacy, instructional quality, and student motivational beliefs: An analysis using multilevel structural equation modeling. *Learning and Instruction*, 66, 101302. <https://doi.org/10.1016/j.learninstruc.2019.101302>
- Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal Of Educational Psychology*, 76(4), 569. <https://doi.org/10.1037/0022-0663.76.4.569>
- Handtke, K., & Bögeholz, S. (2019). Self-efficacy beliefs of interdisciplinary science teaching (SElf-ST) instrument: Drafting a theory-based measurement. *Education Sciences*, 9(4), 247. <https://doi.org/10.3390/educsci9040247>
- Handtke, K., & Bögeholz, S. (2020). Arguments for construct validity of the self-efficacy beliefs of Interdisciplinary Science Teaching (SElf-ST) instrument. *European Journal of Educational Research*, 9(4), 1435-1453. <https://doi.org/10.12973/eu-jer.9.4.1435>
- Hashim, A. M., Aris, S. R. S., & Fook, C. Y. (2022). Assessing gender differences in creative self-efficacy, creative ability and creative environment dimensions among lower secondary school students. *Asian Journal of University Education*, 18(4), 879-893. <https://doi.org/10.24191/ajue.v18i4.19993>
- Henson, R. K. (2001). Teacher self-efficacy: Substantive implications and measurement dilemmas. *Paper presented at the Annual Meeting of the Educational Research Exchange* (College Station, TX, January 26, 2001). <https://shorturl.at/jqEU1>
- Hidayat, R. (2024). Validity of STEM-based modelling instrument for pre-service teachers of mathematics education. *Journal of Institutional Research South East Asia*, 22(1), 154-178
- Hidayat, R., Wan Idris, W. I., Quدراتuddarsi, H., & Abdul Rahman, M. N. (2021). Validation of the mathematical modeling attitude scale for Malaysian mathematics teachers. *Eurasia Journal of Mathematics, Science and Technology Education*, 17(12), 1-16. <https://doi.org/10.29333/ejmste/11375>
- Höltge, L., Ehm, J. H., Hartmann, U., & Hasselhorn, M. (2019). Teachers' self-efficacy beliefs regarding assessment and promotion of school-relevant skills of preschool children. *Early child development and care*, 189(2), 339-351. <https://doi.org/10.1080/03004430.2017.1323888>
- Koe, L. W., Krishnan, R., & Alias, N. E. (2021). The Influence of Self-Efficacy and Individual Entrepreneurial Orientation on Technopreneurial Intention among Bumiputra Undergraduate Students. *Asian Journal of University Education*, 17(4), 490. <https://doi.org/10.24191/ajue.v17i4.16196>
- Lazarides, R., & Warner, L. M. (2020). *Teacher self-efficacy*. Oxford Research Encyclopedia of Education. <https://doi.org/10.1093/acrefore/9780190264093.013.890>
- Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, 18(1), 1-7. <https://doi.org/10.1186/s12874-018-0611-x>
- Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis. *International Journal of Qualitative Methods*, 16(1), 1609406917733847. <https://doi.org/10.1177/1609406917733847>
- O'Neill, A. M. (2023). *Development and validation of an instrument to measure teacher self-efficacy for meeting the needs of military-connected children*. Theses and Dissertations

- Education Sciences. 119. [https://uknowledge.uky.edu/edsc\\_etds/119](https://uknowledge.uky.edu/edsc_etds/119)
- Perera, H. N., & John, J. E. (2020). Teachers' self-efficacy beliefs for teaching math: Relations with teacher and student outcomes. *Contemporary Educational Psychology*, 61, 101842. <https://doi.org/10.1016/j.cedpsych.2020.101842>
- Ratanawongsa, N., Thomas, P. A., Marinopoulos, S. S., Dorman, T., Wilson, L. M., Ashar, B. H., Magaziner, J. L., Miller, R. G., Prokopowicz, G. P., Qayyum, R., & Bass, E. B. (2008). The reported validity and reliability of methods for evaluating continuing medical education: A systematic review. *Academic Medicine*, 83(3), 274–283. <https://doi.org/10.1097/acm.0b013e3181637925>
- Raykov, T., & Grayson, D. (2003). A test for change of composite reliability in scale development. *Multivariate Behavioral Research*, 38(2), 143–159. [https://doi.org/10.1207/s15327906mbr3802\\_1](https://doi.org/10.1207/s15327906mbr3802_1)
- Resnick, B. (2008). *Theory of self-efficacy. Middle range theory for nursing*. <https://doi.org/10.1891/9780826159922.0010>
- Rosli, R., Mokhsein, S. E., & Suppian, Z. (2021). Faktor yang mempengaruhi efikasi sendiri guru dalam melaksanakan pentaksiran bilik darjah di sekolah rendah. *International Journal of Education and Training*, 7, 1-6.
- Saad, M. M. R., Mamat, S., Hidayat, R., & Othman, A. J. (2023). Integrating technology-based instruction and mathematical modelling for STEAM-based language learning: A sociocultural and self-determination theory perspective. *International Journal of Interactive Mobile Technologies*, 17(14), 55–80. <https://doi.org/10.3991/ijim.v17i14.39477>
- Sánchez-Rosas, J., Dyzenchouz, M., Dominguez-Lara, S., & Hayes, A. (2022). Collective teacher self-efficacy scale for elementary school teachers. *International Journal of Instruction*, 15(1), 985-1002. <https://doi.org/10.29333/iji.2022.15156a>
- Siraj, M. A. M. M., Rami, A. A. M., Aziz, N. A. A., & Anuar, M. A. M. (2022). The relationship between school heads' transformational and structural leadership styles towards teachers' teaching motivation. *Asian Journal of University Education*, 18(1), 244-255. <https://doi.org/10.24191/ajue.v18i1.17193>
- Slater, E. V., & Main, S. (2020). A measure of classroom management: Validation of a pre-service teacher self-efficacy scale. *Journal of Education for Teaching*, 46(5), 616–630. <https://doi.org/10.1080/02607476.2020.1770579>
- Smolleck, L. D., Zembal-Saul, C., & Yoder, E. P. (2006). The development and validation of an instrument to measure preservice teachers' self-efficacy in regard to the teaching of science as inquiry. *Journal of Science Teacher Education*, 17(2), 137–163. <https://doi.org/10.1007/s10972-006-9015-6>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., et al. (2018). PRISMA extension for scoping reviews (PRISMA-scr): Checklist and explanation. *Annals of Internal Medicine*, 169(7), Article Number 467e473. <https://doi.org/10.7326/M18-0850>
- Utami, A., Fauziyah, P. Y., Mustadi, A., Hidayat, R., & Rofiki, I. (2024). Bibliometric Analysis of Research Developments on Differentiated Instruction. *European Journal of Educational Research*, 13(3), 1421-1439. <https://doi.org/10.12973/eu-jer.13.3.1421>
- Yang, W., Wu, R., & Li, J. (2021). Development and validation of the STEM Teaching Self-efficacy Scale (STSS) for early childhood teachers. *Current Psychology*, 1-9. <https://doi.org/10.1007/s12144-021-02074-y>
- Yoon, S. Y., Evans, M. G., & Strobel, J. (2014). Validation of the teaching engineering self-efficacy scale for K-12 teachers: A structural equation modeling approach. *Journal of Engineering Education*, 103(3), 463-485. <https://doi.org/10.1002/jee.20049>
- Zakariya, Y. F. (2020). Effects of school climate and teacher self-efficacy on job satisfaction of mostly STEM teachers: A structural multigroup invariance approach. *International Journal of STEM Education*, 7(1), 1–12. <https://doi.org/10.1186/s40594-020-00209-4>
- Zee, M., & Koomen, H. M. Y. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being. *Review of Educational Research*, 86(4), 981–1015. <https://doi.org/10.3102/0034654315626801>

Zhuofan, H., Hidayat, R., & Ayub, A. F. M. (2024). The mediating effect of engagement in the relationship between self-efficacy and perceived learning in the online mathematics environment among Chinese students. *Discover Sustainability*, 5(1). <https://doi.org/10.1007/s43621-024-00586-8>