RESEARCH TRENDS IN CONFIDENCE ASSESSMENT: A SYSTEMATIC LITERATURE REVIEW

HANA FAKHIRA ALMARZUKI[1], KHYRINA AIRIN FARIZA ABU SAMAH[2], LALA SEPTEM RIZA[3], SHARIFALILLAH NORDIN[4]

[1]Lecturer, Universiti Teknologi MARA Shah Alam, College of Computing, Informatics and Mathematics, Malaysia
[2]Senior Lecturer, Universiti Teknologi MARA Cawangan Melaka, College of Computing, Informatics and Mathematics, Malaysia
[3]Professor, Department of Computer Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia
[4]Senior Lecturer, Universiti Teknologi MARA Shah Alam, College of Computing, Informatics and Mathematics, Malaysia
E-mail: [1]hanafakhira@uitm.edu.my*, [2]khyrina783@uitm.edu.my, lala.s.riza@upi.edu, sharifalillah@uitm.edu.my

ABSTRACT

Confidence assessment is vital for validating decisions, managing risks, and improving the quality of knowledge. However, there is a noticeable gap in comprehensive research aimed at investigating, understanding, and interpreting the evolving trends in confidence assessment. In light of this, the study involves conducting a systematic literature review to evaluate the nuances and key findings of previous works, shedding light on the current state of knowledge in this field. Emphasizing the significance of confidence, particularly in learning, is essential for accurately determining a student's level of knowledge. Our systematic review, following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, analyzes 39 studies (2018-2023) from Scopus, Web of Science, and Science Direct. Four primary themes—aims, methods, approaches, and Likert scale types are unfolded into 27 sub-themes, offering a comprehensive view of confidence assessment research trends. Notably, in fields like medicine, confidence assessment is pivotal, with pre-and post-surveys using a 5-point Likert scale being predominant. By synthesizing findings, it informs future research, enhances methods, and contributes to advancing our collective understanding of confidence assessment in knowledge creation.

Keywords: *Confidence Assessment, Systematic Literature Review, Knowledge, Confidence, Research Trend*

1. INTRODUCTION

Confidence is not a sort of knowledge but rather a type of belief that emerges from the subjective observations of an individual's knowledge, which these subjective observations may or may not be accurate [1]. Confidence differs from knowing-to-know (knowledge that is always correct). However, it refers to people's understanding of their knowledge and is a common knowledge concept [1]. Confidence can also be defined as "trust or reliance, a relationship of trust, something confided, and a sense of certainty" [2]. Thus, it is essential to consider confidence in any circumference to achieve the objective, especially involving knowledge.

Assessment refers to the method used to acquire information regarding student ability, knowledge,

and motivation; meanwhile, evaluation is the process of gathering the necessary data to assess if a programme accomplishes its objectives [3]. The "confidence assessment" refers to something that must incorporate the knowledge and skills of all those who participated in the evaluation [4]. Therefore, considering the confidence assessment can enhance the prediction performance to achieve the objective.

Despite the fact that there is a large body of literature on confidence assessment, there has been little effort to analyze these studies systematically, discover trends, and generate prospective themes on the issue. The review procedures, including identification, screening, and eligibility, have not been effectively handled. Traditional literature reviews have several concerns about transparency

and bias. Many authors will select articles in favor of their topic of interest [5].

This study critically examines previous studies on confidence assessment, delving into their methodologies, limitations, and contributions. The analysis considers varying approaches and contextual nuances, providing valuable insights into the strengths and weaknesses of these studies. What sets our work apart is a comprehensive approach driven by the motivation to address a significant gap in comprehensive research on confidence assessment trends. Unlike prior efforts that focused on specific facets, our study adopts a holistic perspective. As a result, future scholars would face a significant challenge in replicating the study, validating the interpretations, or examining the study's comprehensiveness under such a system.

Given this vacuum in the literature, the current study attempts to undertake a Systematic Literature Review (SLR) focusing on the research trend in confidence assessment. Our systematic review methodology, guided by Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, ensures transparency and rigor, contrasting with traditional literature reviews. The empirical results can be justified using this approach to identify gaps and suggest the direction of future research in this field. We set the primary research question that directs our aim: "What are the research trends in confidence assessment?". Furthermore, it is important to know the trend in confidence assessment research to enhance the prediction in terms of confidence. Based on this topic, these SLRs attempt to discuss the following research question:
Question 1: What are the aims of using confidence assessment?
Question 2: What are the methods or approaches that are mostly used when using confidence assessment?

2. METHODS

The SLR aims to locate, search, and synthesize literature systematically related to previous studies or research in a well-organized and transparent process, using replicable procedures throughout each step. Systematic reviews can also be called Meta-narrative reviews or mixed-study reviews [6]. Systematic literature review refers to identifying, evaluating, and interpreting all available research relevant to a specific research question, topic area or phenomenon of interest by using a replicable and detailed methodology [7], [8]. Moreover, identifying

the known and unknown is a time-consuming process. That is the critical reason why systematic literature reviews should be conducted with predefined and transparent methodological steps.

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) were used in this study to establish the SLR on safety culture in mining industries. PRISMA is a very well-known method to conduct SLR in various fields of research, such as safety research [9], [10], social sciences [11], healthcare [12], and business [13]. Besides that, the PRISMA standards offer a well-defined framework for the documentation of systematic reviews, thereby minimizing potential biases and enhancing the reliability and credibility of the research findings [14]. In addition, the utilization of the PRISMA helps authors structure their review process and generate a thorough report, thereby facilitating enhanced comprehension and interpretation of the research findings [15]. These are four main steps for PRISMA: identification, screening, eligibility, and data abstraction and analysis, as shown in Figure 1.

2.1 Identification

Identification is the first stage in conducting a systematic review, which is completed in early December 2023. According to Moher et al. (2009), at least one database must be provided. This review utilized the three main indexed databases: Scopus, Web of Science (WoS), and Science Direct (SD). The systematic review methodology includes three basic processes for choosing multiple eligible publications for this study. It involves thesauruses, dictionaries, and encyclopedias, which involve keyword recognition and searching for connected and related terms.

Consequently, after identifying all relevant phrases, search strings for the Scopus, WoS, and SD databases have been constructed (see Table 1). In the initial step of the systematic review approach, the current study successfully retrieved 461 documents from Scopus, 278 from WoS, and 102 documents from SD. In total, 841 articles were collected for the first phase.

2.2 Screening

The second step, screening, was conducted on the 841 identified articles from the three databases based on the search strings (refer to the identification phase). Screening is the process of including or eliminating articles based on author-established criteria and with the assistance of specialized databases. The screening procedure defined eligibility, inclusion, and exclusion criteria to

identify papers suitable for inclusion in the systematic review procedure. Table 2 displays the inclusion and exclusion criteria established by the authors of the SLR study.
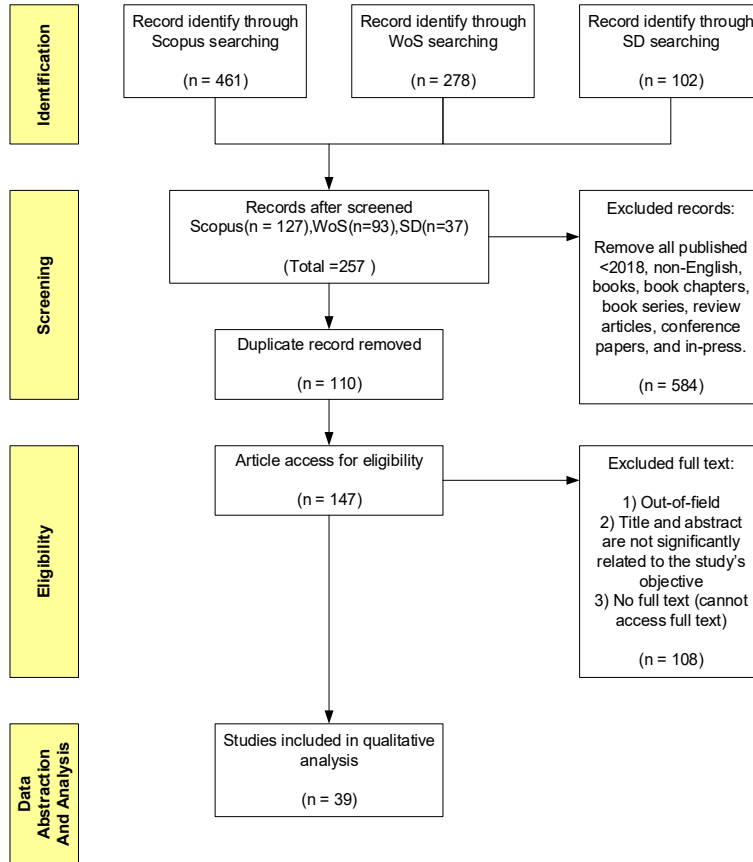

Figure 1: PRISMA Flow Diagram Based on Online Database Identification

Table 1: The Search String

Database	Search String
Scopus	TITLE-ABS-KEY (("confidence-based assessment" OR "confidence based assessment" OR "confidence-based estimate" OR "confidence-based appraisal" OR "confidence level-based assessment" OR "confidence assessment" OR "confidence-based evaluation" OR "confidence based evaluation" OR "confident assessment"))
WoS	topic("confidence-based assessment" OR "confidence based assessment" OR "confidence-based estimate" OR "confidence-based appraisal" OR "confidence level-based assessment" OR "confidence assessment" OR "confidence-based evaluation" OR "confidence based evaluation" OR "confident assessment")
SD	Title, abstract or author-specified keywords ("confidence-based assessment" OR "confidence based assessment" OR "confidence-based estimate" OR "confidence-based appraisal" OR "confidence level-based assessment" OR "confidence assessment" OR "confidence-based evaluation" OR "confidence based evaluation" OR "confident assessment")

Table 2: The Inclusion and Exclusion Criteria

Criterion	Inclusion	Exclusion
Language	English	Non-English
Timeline	2018-2023	<2018

Document Type	Journal (only research articles)	Books, book chapters, book series, review articles, conference papers, and in-press

First, the factor for admission and exclusion was language. All non-English documents were excluded to minimize misunderstandings and translation difficulties. Second, a publication timeline between 2018 and 2023 was chosen based on the total number of retrieved and to-be-reviewed linked publications. The third criterion for inclusion was based on document types. For this criterion, article journals from three databases were picked. Books, book chapters, book series, review articles, conference papers, and in-press were omitted since they did not qualify as primary sources.

In total, 584 articles have been removed, with the balance of 257 articles. Next, the process continued with the elimination of the duplicate documents. From the three databases, 110 identical articles were eliminated, leaving 147 articles for the eligibility process and carried to the next phase.

2.3 Eligibility

The third step is known as eligibility. Eligibility is the ability to manually include or exclude articles based on the authors' particular criteria. Articles that did not fulfill the selected criteria were eliminated from the retrieval process after a thorough review. Then, these articles were screened manually for literature focusing on confident assessment and removed articles due to 1) out-of-field, 2) title and abstract not significantly related to the study's objective, and 3) no full text (cannot access full text). Finally, 108 articles were excluded, and 39 articles were chosen for a comprehensive literature review of confidence assessment research trends.

2.4 Data Abstraction and Analysis

The fourth phase involves the data abstraction and analysis of the data. The remaining articles were evaluated, reviewed, and analyzed. There were 39 selected articles (studies) that are described in detail in this study. The reviews were based on specific studies that matched the study issue. The studies were then extracted to find relevant themes and subthemes for the current study by reading the titles and abstracts, followed by a comprehensive (in-depth) review of the entire article's contents. An integrative review, a form of study synthesizing several research designs, was conducted and completed (qualitative, quantitative, and mixed methods). A thematic analysis was conducted to uncover topics connected to the research trends in confidence assessment studies. The process of identifying and categorizing the major topics, similarities, and differences highlighted and portrayed in the 39 pieces is known as thematic analysis. As Nowell et al. (2017) described, six steps were used in thematic analysis to generate topics for this SLR study. These processes were acceptable for qualitative analysis. The steps consist of:

i. Familiarization with data (understand and analyze the 39 articles)
ii. Generating an initial code (identifying the similarities and differences between the issues discussed in the 39 articles)
iii. Creation of identifiable themes (create suitable themes based on the identified similarities and differences in the 39 articles)
iv. Reviewing themes (ensuring the proposed themes and subthemes are within the main context of each article)
v. Defining and naming themes (four main themes and 27 subthemes were created in this study based on the 30 articles)
vi. Producing a report (in this case, this refers to the SLR study)

3. RESULTS

Table 3 shows the overall review collected 39 articles from nineteen countries, including Belgium, United States, Germany, Ukraine, Arab Saudia, Poland, China, United Kingdom, Malaysia, Korea, New Zealand, Africa, Sweden, Australia, Vietnam, Switzerland, India, France and Brazil. This SLR study generated eight in 2018 and four in 2019. In 2020, seven, 2021, eleven, 2022, five, and 2023, four, respectively. The number of confidence assessment studies from various countries from 2018 to 2023 was determined.

Table 3: Matrix Table on SLR Research Study From 2018 to 2023

Authors	Type of Study	Country	Aims Use Confidence Assessment					Methods									Approaches								Likert Scale Types				
			CKaM	CM	CSD	CKMA	CKCSE	LS	ERSC, P-DW and ISB	C-BEAM	EN	SA	BNN	DNN	CBA	KPCS	1TS	PPS	RCA	Nm	CbMS	CNN	2D	PFA	3pt	4pt	5pt	7pt	10pt
[18]	QN	Germany			✓										✓							✓							
[19]	QN	US	✓					✓										✓									✓		
[20]	QN	US			✓			✓									✓								✓				
[21]	MM	US	✓					✓									✓										✓		
[22]	MM	Belgium	✓															✓											
[23]	QN	US	✓					✓									✓											✓	
[24]	QN	Germany	✓	✓						✓										✓									
[25]	MM	Ukraine	✓					✓										✓								✓			
[26]	QN	Saudi Arabia	✓					✓									✓										✓		
[27]	QN	Poland		✓																									
[28]	QN	China			✓																								
[29]	QN	US		✓				✓									✓											✓	
[30]	QN	US	✓															✓											
[31]	QN	United Kingdom		✓							✓								✓										
[32]	QN	Malaysia	✓					✓									✓		✓						✓				
[33]	QN	Korea		✓										✓									✓						
[34]	QN	US	✓					✓										✓									✓		
[35]	QN	New Zealand		✓				✓									✓									✓			
[36]	MM	Africa	✓					✓										✓									✓		
[37]	QN	Sweden	✓					✓										✓									✓		
[38]	QN	US	✓					✓										✓									✓		
[39]	QN	US		✓				✓										✓							✓				
[40]	QL	US	✓					✓										✓							✓				
[41]	MM	Australia	✓					✓										✓							✓				
[42]	QN	Vietnam			✓																								
[43]	QN	Switzer-land		✓																									
[44]	QN	India		✓						✓											✓								
[45]	MM	Australia	✓					✓										✓									✓		
[46]	QN	France			✓																								
[47]	QN	India	✓					✓										✓									✓		
[48]	MM	United Kingdom		✓				✓									✓										✓		
[49]	QN	US	✓					✓									✓										✓		
[50]	QN	China			✓																								
[51]	QN	US					✓	✓										✓											✓
[52]	QN	US	✓					✓										✓									✓		
[53]	MM	United Kingdom				✓		✓									✓												✓
[54]	QN	Brazil		✓												✓	✓												
[55]	QN	Germany		✓								✓																	
[56]	QN	China		✓										✓								✓							

Theme 1: Aims Use Confidence Assessment

CKaM = Confidence in Knowledge of Medical
CM = Confidence in Method
CSD = Confidence System Development
CKMA = Confidence Knowledge in Mathematics Assessment
CKCSE = Confidence in Knowledge Car Seat Education

Theme 2: Methods

LS = Likert Scale
ERSC, P-DW and ISB = Explicit Reports of Subjective Confidence, Post-Decision Wagering and Info-Seeking Behavior
C-BEAM = Confidence-Based Evaluation Approach for MCQ
EN = Entropy
SA = SMILES Augmentation
BNN = Bayesian Neural Network
DNN = Deep Neural Network model
CBA = Confidence-Based Assessment
KPCS = Karitane Parenting Confidence Scale

Theme 3: Approaches

1TS = One Time Survey
PPS = Pre And Post-Survey
RCA = Reverse-Correlation Analysis
NM = Naïve Model
CbMS = Confidence-based Marking Scheme
CNN = CNN Architecture
2D = Two Dimension
PFA = Process-Focused Assessment

Theme 4: Likert Scale Types

3pt = 3-point
4pt = 4-point
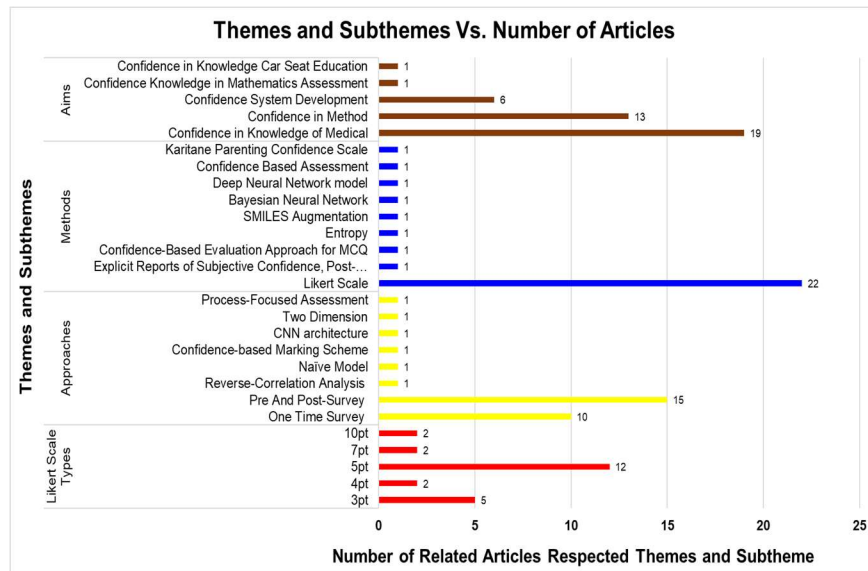5pt = 5-point
7pt = 7-point
10pt = 10-point

Figure 2: Themes and Subthemes Distribution Based on Collected Articles

The United States was a leading country for published papers, with thirteen articles and studies on confidence assessment, followed by China, the United Kingdom and Germany, with three studies: Australia and India had two studies each. One paper is from Brazil, France, Sweden, Africa, New Zealand, Korea, Malaysia, Poland, Saudi Arabia, Ukraine, Belgium, Vietnam and Switzerland. Furthermore, for the types of studies in this research, the majority of the articles were quantitative (30 articles). In comparison, the mixed method had both quantitative and qualitative (seven articles) and qualitative (one article).

Four major themes and twenty-seven subthemes were developed through the thematic analysis adaptation. It consists of the aims of using confidence assessment represented as aims (5 subthemes), methods when using confidence assessment represented as methods (9 subthemes), approaches when using confidence assessment represented as approaches (8 subthemes), and the type of Likert scale represented as Likert scale types (5 subthemes). The distribution is presented in statistical form, including the number of frequencies at which the 27 subthemes appeared in the articles.

Based on the thematic analysis, Figure 2 depicts the distribution of articles by main themes and subthemes. The highest reported numbers are emerging under the themes of aim using confidence assessment involving confidence in knowledge of medical (19 studies). As for methods, the Likert scale ranked the highest (23 studies), and approaches when using confidence assessment were the pre- and post-survey (15 studies) and Likert scale types involving 5 Likert scales (12 studies).

Figure 3 represents the summary percentages of the four major themes among the twenty-seven subthemes. The aims using confidence assessment accounted for 19% (5 out of 27 articles for subthemes), methods 33% (9 out of 27 articles for subthemes), approaches 30% (8 out of 27 articles for subthemes), and types of Likert scale 19% (5 out of 27 articles for subthemes). Thus, it can be concluded that most studies in confidence assessment focus on the method used in measuring effectiveness. The Likert scale is the most discussed method, with 23 articles and 9 subthemes.
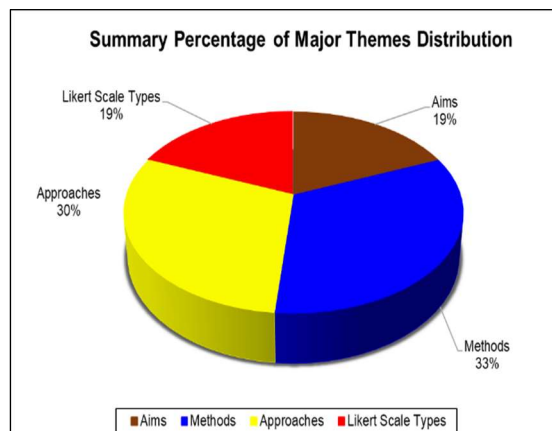

Figure 3: Summary Percentage of Themes Distribution

4. DISCUSSION

The main purpose of performing a systematic review is to analyze the factors involved in using confidence assessments. It is crucial to have confidence, especially in learning, so that the knowledge of the student can indeed be determined. The current analysis found 39 papers that functioned as four major themes: 1) aims of using confidence assessment, 2) methods when using confidence assessment, 3) approaches when using confidence assessment, and 4) Likert scale types used for confidence assessment.

4.1 Aims of Using Confidence Assessment

Confidence in knowledge of medical (CKaM), confidence in the method (CM) used, confidence in system development (CSD), confidence in knowledge about mathematics assessment (CKMA), and confidence in knowledge about car seat education (CKCSE) were identified as the aims of the usage of confidence assessment. As shown in Figure 4, confidence in knowledge of medical has the highest percentage of studies (48%). Confidence assessment has been used in higher education, particularly in medical, where it is critically important to discourage guessing in life-and-death matters. This is due to the implementation of confidence assessment in learning, especially in the medical field, where it is crucial to discourage guessing in life-and-death situations [53]. It is important to stress that the aim of using a confidence assessment is not only to determine the student's confidence but also to help enable more effective learning in the future. It is parallel to the fact that confidence conveys the perception that an individual is capable of achieving a desired objective [26].
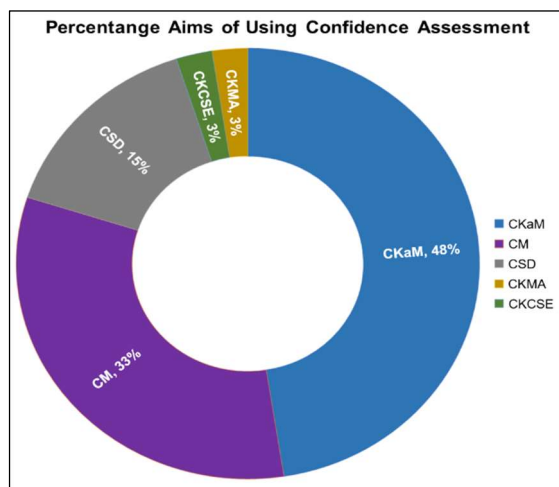

Figure 4: Percentage Aims of Using Confidence Assessment

Besides that, the aim of knowing the confidence in the methods used is to study the relationship between the estimation error and the uncertainty [56]. It is made by analyzing the confidence and prediction error [55]. It is important to examine the uncertainty in the input data to describe how reliable the data are for a given purpose [43]. Consequently, it is crucial to assess the confidence and uncertainty in addition to the prediction objective.

Confidence in system development aims to ensure that the development of the system process is effectively developed since it is a significant challenge, especially when it is related to safety cases [46]. Moreover, it is important to analyze the credibility of the development of the system, such as the method for calculating the confidence value and the method for predicting the confidence interval, since this will be the validation of the guidance and the control system. Therefore, analyzing this confidence in system development will improve the credibility of the system.

4.2 Methods When Using Confidence Assessment

Likert Scale (LS), Explicit Reports of Subjective Confidence (ERSC), Post-Decision Wagering (PDW) and Info-Seeking Behavior (ISB), Confidence-Based Evaluation Approach for MCQ (C-BEAM), Entropy (EN), SMILES Augmentation (SA), Bayesian Neural Network (BNN), Deep Neural Network (DNN) model, Confidence Based Assessment (CBA), and Karitane Parenting Confidence Scale (KPCS) were the methods used when using confidence assessment. Mostly, the Likert scale method was used to predict the knowledge of the students in the medical field (16 studies). Each aim consists of a Likert Scale as the method, except for confidence system development (see Table 4). As for confidence in the methods used, most have their own method and approach in order to know the confidence of the method used (7 studies). Meanwhile, confidence in the system's development is used, focusing on the framework, design, and methodology [28], [42], [50], [46].

The majority of the methods used when using confidence assessments are to predict knowledge. Knowledge shortcomings or uncertainty are part of the knowledge gap that must be considered [57]. Therefore, it is important to evaluate one's self-perceived knowledge after performing a task or test to determine how well it performed after completing a task. Based on the SLR

study, the method of evaluation to evaluate knowledge is the Likert scale. In educational and social sciences research, the Likert scale is one of the most fundamental and widely employed psychometric instruments [58].

4.3 Approaches When Using Confidence Assessment

Most of the methods that were utilized each took their own unique approach, such as the Likert scale, which is based on surveys and may be further subdivided into pre- and post-surveys as well as one-time surveys. This one-time survey was conducted at the conclusion of the learning experience. In contrast, the pre- and post-learning surveys were conducted, respectively, before and after the learning experience [49], [34]. Other than the Likert Scale, Explicit Reports of Subjective Confidence, Post-Decision Wagering and Info-Seeking

Behavior, Confidence-Based Evaluation Approach for MCQ, Entropy, SMILES Augmentation, Bayesian Neural Network, Deep Neural Network Model, Two Dimension, and Karitane Parenting Confidence Scale. All of these measures can be found in the article.

According to Table 4, an approach known as Reverse-Correlation analysis is used when Explicit Reports of Subjective Confidence, Post-Decision Wagering and Info-Seeking Behavior activity are collected. Next, the approach of the C-BEAM is a confidence-based marking scheme; entropy is a Naive Model; Bayesian Neural Network architecture is CNN architecture; the Deep Neural Network model is a Process-Focused Assessment (PFA); and the Karitane Parenting Confidence Scale is a one-time survey.

Table 4: The Distribution of Methods and Approaches based on Aims Using Confidence Assessment

Aims		Methods	Total	Approaches	Total
CKaM	19	Likert Scale	16	One-time survey	5
				Pre and post-survey	11
		No Likert Scale	2	No Likert	2
		ERSC, P-DW and ISB	1	RCA	1
CM	13	Likert Scale	3	1 Time Survey	3
		Likert Scale	1	Pre and post-survey	1
		Explicit reports of subjective confidence, post-decision wagering and info-seeking behavior	1	Reverse-correlation analysis	1
		Confidence-based evaluation approach for MCQs (C-BEAM)	1	Confidence-based marking scheme	1
		Entropy	1	Naïve model	1
		SMILES augmentation	1		1
		Bayesian neural network	1	CNN architecture	1
		Deep neural network model	1	Process-Focused Assessment (PFA)	1
		Karitane Parenting Confidence Scale	1	One-time survey	1
CSD	6	No specific methods	2	No specific approaches	2
CSD	6	Confidence Based Assessment	1	Two dimension	1
CKSE	1	Likert Scale	1	1 Time Survey	1

4.4 Likert Scale Types Used for Confidence Assessment

The majority of the method used to predict confidence in student knowledge is the Likert scale method. In this Likert scale method, there are various types of scales used for confidence assessment. Based on the SLR study, there are a 3-point Likert scale, a 4-point Likert scale, a 5-point Likert scale, a 7-point Likert scale and a 10-point

Likert scale, as shown in Figure 5. The most commonly used type of scale is the 5-point Likert scale. The 5-point Likert scale is used to differentiate the level of confidence in 5 terms: 1 = Not at all, 2 = A little bit, 3 = Comfortable, 4 = Very confident, and 5 = Extremely confident [34]. Basically, it went from nothing at all to very/extremely confident.

Besides that, the confidence assessment items were rated on a 3-point Likert scale, where 1 = Non-confident, 2 = Not sure, and 3 = Confident [32]. Based on Patel et al. (2020), this 3-point Likert scale can also be utilized as the capacity to accomplish the task in question as 1 = Uncertain, 2 = Sometimes confident, or 3 = Most of the time confident or estimate by using frequencies of >75% = 3, 25% to 75% = 2, and <25% = 1. The rest of the types of the Likert scale, which is a 7-point Likert scale and a 10-point Likert scale, were identified in two studies each, while one study was for a 4-point Likert scale.
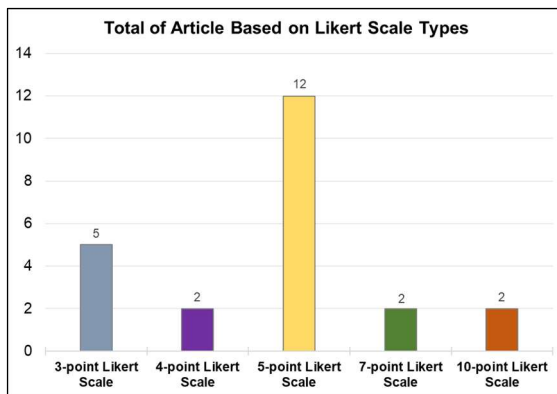
Figure 5: Number of Articles based on Types of Likert Scale

4.5 Research Gaps and Recommendations

The results of this study have significant implications for developing educational guidelines, rules, or procedures that specifically target confidence assessments. They are particularly relevant for educators, students, curriculum developers, learning practitioners, and policymakers, helping to address potential shortcomings that might hinder efficient confidence assessment, create conducive learning environments, and enhance future knowledge quality. Despite meticulous attention to detail throughout the execution of this SLR, certain limitations persist. Based on the findings of the present study, future researchers are advised to investigate the following gaps:

- The search technique focused on indexed journals, which are Scopus, WoS and SD. This SLR excludes non-indexed publications because they do not meet the predetermined criteria for inclusion. Researchers may expand the existing study use of numerous databases or search engines, such as Google Scholar, to gather more information.
- Given the widespread interest in confidence assessment, there might be more empirical studies in other languages that could confirm,

clarify, or challenge this SLR's findings. Only English-language publications that are available through a university library system and have undergone peer review.

- Along with the methods and results of the studies, there was an evaluation process that improved the accuracy of the evaluation part. However, the work that went into combining results from both qualitative and quantitative data analyses (which were only done in two studies) might not have allowed all methodological issues to be fully addressed when the results were put together.
- Possible approaches involve consulting with subject matter specialists, conducting citation tracking, performing reference searches, and utilizing the snowballing technique.

5. CONCLUSION

The primary objective of this study is to conduct a SLR focused on researching trends in confidence assessment from 2018 to 2023. Employing the PRISMA approach and employing thematic analysis, we systematically reviewed 40 meticulously selected publications sourced from Scopus, WoS, and SD. The study categorically addressed four major themes: aims, methods, approaches, and types of Likert scales. Within these, we identified twenty-seven subthemes, encompassing the objectives of confidence assessment (19%), methodologies employed (33%), approaches utilized (30%), and Likert scale types applied (19%). The SLR revealed that Likert scales and surveys were the predominant methods and strategies employed across the selected studies. Notably, most of the research focused on confidence assessment as a means of gauging confidence in knowledge and skills, with a significant emphasis on the medical field, often utilizing one-time surveys and pre- and post-surveys. Recognizing the crucial role of confidence assessment in predicting student mastery levels in knowledge and skills, this study underscores its importance in gauging genuine comprehension of learning. In conclusion, the findings of this SLR hold potential significance for educators, curriculum developers, and learning facilitators, offering insights to enhance understanding and address issues associated with confidence assessment practices. The implications extend to students, educational policymakers, and governments, contributing to the refinement of confidence assessment practices and, consequently, fostering improved learning outcomes in the future.

REFERENCES

[1] D. Kaleci and E. Akleman, "World Journal on Educational Technology: Current Issues," *World Journal on Educational Technology*, vol. 11, no. 1, pp. 104–115, 2018.

[2] K. A. White, "Self-Confidence: A Concept Analysis," *Nursing Forum*, vol. 44, no. 2, pp. 103–114, 2009.

[3] I. Tosuncuoglu, "Importance of Assessment in ELT," *Journal of Education and Training Studies*, vol. 6, no. 9, p. 163, 2018, doi: 10.11114/jets.v6i9.3443.

[4] H. M. Paula and M. S. Ogaz, "Confidence Level Assessment in Enterprise Risk Management : Case Study with Focus on Oil & Gas Operational Incidents," in *22nd Annual International Symposium*, 2019, pp. 1–22.

[5] H. A. M. Shaffril, S. F. Samsuddin, and A. A. Samah, "The ABC of Systematic Literature Review: The Basic Methodological Guidance for Beginners," *Quality and Quantity*, vol. 55, no. 4, pp. 1319–1346, 2021, doi: 10.1007/s11135-020-01059-6.

[6] G. Wong, T. Greenhalgh, G. Westhorp, J. Buckingham, and R. Pawson, "RAMESES Publication Standards: Meta-Narrative Reviews," *Journal of Advanced Nursing*, vol. 69, no. 5, pp. 987–1004, 2013, doi: 10.1111/jan.12092.

[7] Barbara Kitchenham, "Procedures for Performing Systematic Reviews," 2004.

[8] D. J. Cook, D. L. Sackett, and W. O. Spitzer, "Methodologic Guidelines for Systematic Reviews of Randomized Control Trials in Health Care from the Potsdam Consultation on Meta-Analysis," *Journal of Clinical Epidemiology*, vol. 48, no. 1, pp. 167–171, 1995, doi: 10.1016/0895-4356(94)00172-M.

[9] E. Adaku, N. A. Ankrah, and I. E. Ndekugri, "Design for Occupational Safety and Health: A Theoretical Framework for Organisational Capability," *Safety Science*, vol. 133, 2021, doi: 10.1016/j.ssci.2020.105005.

[10] W. Nyoni, M. Pillay, M. Rubin, and M. Jefferies, *Organizational Factors, Residual Risk Management and Accident Causation in the Mining Industry: A Systematic Literature Review*, vol. 791, no. July 2018. Springer International Publishing, 2019. doi: 10.1007/978-3-319-94589-7_2.

[11] H. A. Mohamed Shaffril, A. A. Samah, S. F. Samsuddin, and Z. Ali, "Mirror-Mirror on The Wall, What Climate Change Adaptation Strategies are Practiced by The Asian's Fishermen Of All?," *Journal of Cleaner Production*, vol. 232, pp. 104–117, 2019, doi: 10.1016/j.jclepro.2019.05.262.

[12] S. Danielli, T. Coffey, H. Ashrafian, and A. Darzi, "Systematic Review Into City Interventions to Address Obesity," *Eclinicalmedicine*, vol. 32, 2021, doi: 10.1016/j.eclinm.2020.100710.

[13] M. Cubric, "Drivers, Barriers And Social Considerations for AI Adoption in Business and Management: A Tertiary Study," *Technology in Society*, vol. 62, 2020, doi: 10.1016/j.techsoc.2020.101257.

[14] M. J. Page *et al.*, "PRISMA 2020 Explanation And Elaboration: Updated Guidance and Exemplars for Reporting Systematic Reviews," *The BMJ*, vol. 372, 2021, doi: 10.1136/bmj.n160.

[15] A. A. Selcuk, "A Guide for Systematic Reviews: PRISMA.," *Turkish Archives of Otorhinolaryngology*, vol. 57, no. 1, pp. 57–58, 2019, doi: 10.5152/tao.2019.4058.

[16] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Academia and Clinic Annals of Internal Medicine Preferred Reporting Items for Systematic Reviews and Meta-Analyses," *Annals of Internal Medicine*, vol. 151, no. 4, pp. 264–269, 2009.

[17] L. S. Nowell, J. M. Norris, D. E. White, and N. J. Moules, "Thematic Analysis: Striving to Meet The Trustworthiness Criteria," *International Journal of Qualitative Methods*, vol. 16, no. 1, pp. 1–13, 2017, doi: 10.1177/1609406917733847.

[18] F. Lyrath, C. Stechert, and S. I. U. Ahmed, "Application of Augmented Reality (AR) in The Laboratory for Experimental Physics," *Procedia CIRP*, vol. 119, pp. 170–175, 2023, doi: 10.1016/j.procir.2023.03.089.

[19] B. P. Elliott, A. Berglund, R. Markert, and K. Burtson, "Confidence and Utilization are Poorly Associated with Point-Of-Care Ultrasound Competency Among Internal Medicine Trainees," *Military medicine*, vol. 188, no. 6, pp. 316–321, 2023, doi: 10.1093/milmed/usad127.

[20] S. Biswas, J. Follum, and J. H. Eto, "Confidence Assessment for Regional Forced Oscillation Source Localization: Formulation

and Field Validation," *IEEE Transactions on Power Delivery*, vol. 38, no. 6, pp. 3739–3748, 2023, doi: 10.1109/TPWRD.2023.3284418.

[21] D. Owerko, K. Ryan, E. Cabacungan, K. Yan, and K. Saudek, "Neonatal Hyperbilirubinemia: Assessing Variation in Knowledge and Practice," *PLoS ONE*, vol. 18, no. 2, pp. 1–9, 2023, doi: 10.1371/journal.pone.0282413.

[22] M. van de Wiel, K. Bombeke, and A. Janssens, "Communication Skills Training in Advance Care Planning: A Survey Among Medical Students At The University Of Antwerp," *BMC Palliative Care*, vol. 21, no. 1, pp. 1–8, 2022, doi: 10.1186/s12904-022-01042-y.

[23] P. D. Sonenthal *et al.*, "Applying the WHO-ICRC BEC course to Train Emergency and Inpatient Healthcare Workers in Sierra Leone Early in The COVID-19 Outbreak," *BMC Health Services Research*, vol. 22, no. 1, pp. 1–11, 2022, doi: 10.1186/s12913-022-07556-8.

[24] L. Weise, S. D. Forster, and S. Gauggel, "Reverse-Correlation Reveals Internal Error-Corrections During Information-Seeking," *Metacognition and Learning*, vol. 17, no. 2. pp. 321–335, 2022. doi: 10.1007/s11409-021-09286-4.

[25] S. M. Kivlehan *et al.*, "Evaluation of Change in Emergency Care Knowledge and Skills Among Front-Line Healthcare Providers in Ukraine with The Basic Emergency Care Course: A Pretest/Post-Test Study," *BMJ Open*, vol. 12, no. 6, 2022, doi: 10.1136/bmjopen-2021-050871.

[26] D. E. Meisha and R. A. Al-dabbagh, "Self-Confidence as A Predictor Of Senior Dental Student Academic Success," *Journal of Dental Education*, vol. 85, no. 9, pp. 1497–1503, 2021, doi: 10.1002/jdd.12617.

[27] M. Bukowski, J. Kurek, I. Antoniuk, and A. Jegorowa, "Decision Confidence Assessment in Multi-Class Classification," *Sensors*, vol. 21, no. 11, pp. 1–15, 2021, doi: 10.3390/s21113834.

[28] Y. Hu, Y. Wang, S. Wang, and X. Zhao, "Fusion Key Frame Image Confidence Assessment of The Medical Service Robot Whole Scene Reconstruction," *Journal of Imaging Science and Technology*, vol. 65, no. 3, pp. 1–9, 2021, doi: 10.2352/J.ImagingSci.Technol.2021.65.3.030409.

[29] L. M. L. Agler, K. Noguchi, and L. K. Alfsen, "Personality Traits as Predictors of Reading Comprehension and Metacomprehension Accuracy," *Current Psychology*, vol. 40, no.

[30] M. Hughes *et al.*, "Adaptive Change in Simulation Education: Comparison of Effectiveness of A Communication Skill Curriculum on Death Notification Using In-Person Methods Versus A Digital Communication Platform," *AEM Education and Training*, vol. 5, no. 3, 2021, doi: 10.1002/aet2.10610.

[31] G. N. Tornetta, "Entropy Methods for The Confidence Assessment of Probabilistic Classification Models," *Statistica*, vol. 81, no. 4, pp. 383–398, 2021, doi: 10.6092/issn.1973-2201/11479.

[32] J. A. Dujaili, A. Blebil, and D. Chong, "Assessment of The Nutritional Knowledge of Undergraduate Pharmacy Students: The Need to Advance Nutrition Education in The Training Of Pharmacy Students," *Pharmacy Education*, vol. 21, no. 1, pp. 310–314, 2021, doi: 10.46542/pe.2021.211.310314.

[33] H.-J. Lee and D. Lee, "Study Of Process-Focused Assessment Using An Algorithm for Facial Expression Recognition Based on A Deep Neural Network Model," *Electronics (Switzerland)*, vol. 10, no. 1, pp. 1–14, 2021, doi: 10.3390/electronics10010054.

[34] R. Kohli, G. Arora, A. F. Blanc, E. Pham, and P. Gubrud-Howe, "Oral Health Clinical Training and Dental Referral Program for Nurses: An Interprofessional Collaborative Project," *Geriatric Nursing*, vol. 42, no. 4, pp. 880–886, 2021, doi: 10.1016/j.gerinurse.2021.04.015.

[35] J. Cowlyn, B. M. Kennedy, D. M. Gravley, and S. J. Cronin, "A Confidence-Based Assessment Method for Distinguishing Pyroclastic Density Current Deposits from Other Volcaniclastic Units," *Frontiers in Earth Science*, vol. 8, no. November, pp. 1–14, 2020, doi: 10.3389/feart.2020.581195.

[36] J. K. Maddry *et al.*, "Development and Evaluation of An Abbreviated Extracorporeal Membrane Oxygenation (ECMO) Course for Nonsurgical Physicians and Nurses," *AEM Education and Training*, vol. 4, no. 4, pp. 347–358, 2020, doi: 10.1002/aet2.10447.

[37] K. Dietrich, E. Hutchinson, and M. D. Lu, "Advanced Primary Care Orthopedics: Bridging The Gap in Musculoskeletal Education," *Family Medicine*, vol. 52, no. 6, pp. 444–447, 2020, doi: 10.22454/FamMed.2020.362716.

[38] J. Berz, K. Donovan, and M. Eyllon, "An

Interprofessional Nutrition Education Session for Senior Medical Students on Evidence-Based Diet Patterns and Practical Nutrition Tips," *MedEdPORTAL : the journal of teaching and learning resources*, vol. 16, p. 10876, 2020, doi: 10.15766/mep_2374-8265.10876.

[39] A. A. Patel *et al.*, "Development and Evaluation of A Pediatric Epilepsy Training Program for First Level Providers in Zambia," *Global Pediatric Health*, vol. 7. 2020. doi: 10.1177/2333794X20968718.

[40] L. M. Caldas, A. T. Matulewicz, R. A. Koenig, M. Hindle, and K. L. Donohoe, "Using Immersive Simulation to Engage Student Learners in A Nonsterile Compounding Skills Laboratory Course," *Currents in Pharmacy Teaching and Learning*, vol. 12, no. 3, pp. 313–319, 2020, doi: 10.1016/j.cptl.2019.12.016.

[41] C. Scuderi and T. Pain, "Using Confidence Assessment to Explore Physiotherapists, Dieticians, Speech Pathologists and Occupational Therapists Knowledge of Medicines," *Journal of Interprofessional Care*, vol. 33, no. 5, pp. 593–597, 2019, doi: 10.1080/13561820.2018.1551860.

[42] L. H. Nguyen and Q. H. Nguyen, "Confidence-Based Assessment of Safety of Irrigation Reservoir in Vietnam," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 10, pp. 2884–2893, 2019, doi: 10.35940/ijitee.J9616.0881019.

[43] A. M. Kaskela *et al.*, "Picking Up The Pieces—Harmonising and Collating Seabed Substrate Data for European Maritime Areas," *Geosciences (Switzerland)*, vol. 9, no. 2, pp. 1–18, 2019, doi: 10.3390/geosciences9020084.

[44] A. Aleem and M. M. Gore, "C-BEAM: A Confidence-Based Evaluation of MCQs for Providing Feedback to Instructors," *Computer Applications in Engineering Education*, vol. 27, no. 1, pp. 112–127, 2019, doi: 10.1002/cae.22061.

[45] H. L. Hattingh, D. Robinson, and A. Kelly, "Evaluation of A Simulation-Based Hospital Pharmacy Training Package for Pharmacy Students," *International Journal of Educational Technology in Higher Education*, vol. 15, no. 1, 2018, doi: 10.1186/s41239-018-0120-3.

[46] R. Wang, J. Guiochet, G. Motet, and W. Schön, "Modelling Confidence in Railway Safety Case," *Safety science*, vol. 110, pp. 286–299, 2018.

[47] A. S. Wajekar, Sweta V Salgaonkar, Indrani H Chincholi, and A. N. Shetty, "Impact of Basic Medical Writing Workshop on Case Report Writing by Post-Graduate Anaesthesia Trainees: A Pilot Study," *Indian Journal of Anaesthesia*, vol. 62, no. 7, pp. 502–508, 2018, doi: 10.4103/ija.IJA_98_18.

[48] J. Osgerby, P. Jennings, and A. Bonathan, "Do Students See The Benefits? An Exploratory Study of Undergraduate Accounting Students' Perceptions of A Programme Focussed Assessment," *International Journal of Management Education*, vol. 16, no. 2, pp. 327–339, 2018, doi: 10.1016/j.ijme.2018.04.006.

[49] A. L. Marshall *et al.*, "Internal Medicine Trainees' Knowledge and Confidence in Using The American Society of Hematology Choosing Wisely Guidelines In Hemostasis, Thrombosis, and Non-Malignant Hematology," *PLoS ONE*, vol. 13, no. 5, pp. 1–10, 2018, doi: 10.1371/journal.pone.0197414.

[50] H. Wang, X. Dong, Q. Li, and Z. Ren, "Confidence Assessment and Interval Prediction for Multi-Input Model Via Grey System Theory," *Grey Systems*, vol. 8, no. 1, pp. 69–83, 2018, doi: 10.1108/GS-07-2017-0024.

[51] E. Kuroiwa, R. L. Ragar, C. S. Langlais, A. Baker, M. E. Linnaus, and D. M. Notrica, "Car Seat Education: A Randomized Controlled Trial of Teaching Methods," *Injury*, vol. 49, no. 7, pp. 1272–1277, 2018, doi: 10.1016/j.injury.2018.05.003.

[52] A. S. Hafer, W. B. Sweeney, A. L. Battista, H. S. Meyer, and B. R. Franklin, "Development and Implementation of Urologic Care Army/Air Force/Navy Provider Education, A Urologic Emergency Simulation Curriculum," *Military Medicine*, vol. 00, pp. 1–7, 2022, doi: 10.1093/milmed/usac003.

[53] C. Foster, "Implementing Confidence Assessment in Low-Stakes, Formative Mathematics Assessments," *International Journal of Science and Mathematics Education*, vol. 20, no. 7, pp. 1411–1429, 2021, doi: 10.1007/s10763-021-10207-9.

[54] L. W. Pereira, J. R. Bernardi, S. de Matos, C. H. d. Silva, M. Z. Goldani, and V. L. Bosa, "Cross-Cultural Adaptation and Validation of The Karitane Parenting Confidence Scale of Maternal Confidence Assessment for Use In Brazil," *Jornal de Pediatria*, vol. 94, no. 2, pp. 192–199, 2018, doi: 10.1016/j.jped.2017.05.011.

[55] T. B. Kimber, M. Gagnebin, and A. Volkamer, "Maxsmi: Maximizing Molecular Property Prediction Performance With Confidence Estimation Using SMILES Augmentation and Deep Learning," *Artificial Intelligence in the Life Sciences*, vol. 1, p. 100014, 2021, doi: 10.1016/j.ailsci.2021.100014.

[56] W. Xue, T. Guo, and D. Ni, "Left Ventricle Quantification With Sample-Level Confidence Estimation Via Bayesian Neural Network," *Computerized Medical Imaging and Graphics*, vol. 84, p. 101753, 2020, doi: 10.1016/j.compmedimag.2020.101753.

[57] M. Versteeg and P. Steendijk, "Putting Post-Decision Wagering to The Test: A Measure Of Self-Perceived Knowledge in Basic Sciences?," *Perspectives on Medical Education*, vol. 8, no. 1, pp. 9–16, 2019, doi: 10.1007/s40037-019-0495-4.

[58] A. Joshi, S. Kale, S. Chandel, and D. Pal, "Likert Scale: Explored and Explained," *British Journal of Applied Science & Technology*, vol. 7, no. 4, pp. 396–403, 2015, doi: 10.9734/bjast/2015/14975.