International Journal of Assessment Tools in Education
2025, Vol. 12, No. 3, 600–628
https://doi.org/10.21449/ijate.1523590
journal homepage: https://dergipark.org.tr/en/pub/ijate
Research Article

Development and validation of the Tamil Inclusive Classroom Management Self-Efficacy (TICMSE) scale

Yogaranee Sakthivel [1*]

[1]University of Colombo, Faculty of Education, Department of Educational Psychology, Sri Lanka

ARTICLE HISTORY
Received: July 28, 2024
Accepted: Dec. 17, 2024

Keywords:
Classroom management,
Self-Efficacy,
Preservice teachers,
Inclusive education,
Scale development,
Validation.

Abstract: Assessing preservice teachers' inclusive classroom management self-efficacy (ICMSE) is critical for effectively implementing classroom management practices. Although the literature provides well-established instruments, many do not cover all domains of ICMSE. Above all, there is a notable absence of instruments assessing this multidimensional construct in Tamil, one of the native languages of Sri Lanka. This paper comprehensively reports the methodological aspects of developing and validating the Tamil Inclusive Classroom Management Self-Efficacy (TICMSE) Scale. The main purpose of this study is to investigate whether the TICMSE Scale effectively captures the multidimensional nature of the construct of ICMSE. Specifically, the study examines whether the author-generated items accurately represent the hypothesized, a priori four-factor structure of the scale. A convenience sample of 200 preservice teachers enrolled in teacher education institutions in Sri Lanka participated in an online survey. Exploratory Factor Analysis was conducted to extract the data, resulting in a four-factor solution. Confirmatory Factor Analysis further confirmed the proposed structure. The psychometric properties of the TICMSE scale, including composite reliability, internal consistency, and convergent and discriminant validity, were established. These findings warrant the ongoing utilization of this scale as a reliable and factorially valid instrument for objectively measuring different dimensions of ICMSE, especially among Tamil preservice teachers in Sri Lanka.

1. INTRODUCTION

An emerging consensus regarding children with special educational needs (SEN) is that they should be educated in the least restrictive environment, regardless of their diverse backgrounds, abilities, and needs, alongside typically developing peers. This approach ensures equitable access to learning opportunities for all, enabling the fullest academic progress and social integration (UNESCO, 1994). In practice, educators strive to achieve the aim of inclusive education (IE) by accommodating children with SEN in regular classrooms for as much of the day as possible and effectively meeting their diverse needs, thereby increasing their presence, participation, and achievement in learning while reducing exclusion within and from education. To this end, they must master various skills, including instruction and classroom management (CM), two critical pedagogical skills deeply intertwined to ensure effective implementation of

*CONTACT: Yogaranee SAKTHIVEL yogaranee@edpsy.cmb.ac.lk University of Colombo, Faculty of Education, Department of Educational Psychology, Sri Lanka

e-ISSN: 2148-7456

IE (Ainscow & Miles, 2009; O'Neill, 2015; UNESCO, 2008; UNESCO, 2019). Providing an optimal learning environment through effective CM is one of the main tasks of teachers to foster students' overall well-being and academic and social-emotional development (Garrote *et al*., 2020). However, to what extent this task is being accomplished remains an ever-emerging question (Marzano *et al*., 2003).

The critical role that CM plays in promoting learning, minimizing disruptions and distractions in the classroom, and fostering social-emotional development in students has been well-documented in the literature for more than decades (Brophy, 2006; Doyle, 1986; Emmer *et al*., 1980; Oliver & Reschly, 2007; Oliver *et al*., 2011). However, extant literature finds that CM remains a significant source of concern for school systems, and indeed, assistance with CM is the most frequent request made by teachers (Alter *et al*., 2013; Alter & Haydon, 2017; Oliver *et al*., 2011). Teachers placed CM at the top of the list of challenges they face and of tasks that generate anxiety when dealing with an increased number of students with diverse needs. As a result, they are reported to have concerns about implementing CM in inclusive classrooms (Dicke *et al*., 2014; Eisenman *et al*., 2015; Milner & Tenore, 2010; Oliver & Reschly, 2007). Most importantly, preservice teachers appeared to be more conscious of students' inappropriate behavior and their ability to manage classrooms with such behavior (Page & Jones, 2018; Subban & Round, 2015; Yogaranee, 2024). "If teachers do not react adequately to students when their behavior is disruptive, instructional time is lost for all students. To reach instructional goals, teachers must deal adequately with disruptive behavior in the classroom" (Brouwers & Tomic, 2000, p. 242).

The concept of CM, in general, encompasses a set of strategies, practices, and procedures teachers adopt to ensure orderliness and smooth conduct of teaching and learning in the classroom. However, its conceptualization differs in different fields of study, including management and psychology. From a management perspective, CM refers to teachers' strategies to organize students, space, time, and materials to create an optimal work environment for students to learn and achieve (Wong & Wong, 2009). The field of psychology conceptualizes CM by adding student behaviors to the top of managing resources as conceptualized in the management field (Marzano *et al*., 2003).

Traditionally, student disruptive behavior has been a central focus in many discussions of CM, with researchers often attributing misconduct to individual student characteristics, typically as a consequence of teachers' inadequate CM skills (Kaplan *et al*., 2002). However, Doyle (1986) asserted that when teachers are preoccupied with student misbehavior, they are more likely to make reactive decisions such as reprimands, punishments, token economies, and contingencies while ignoring organizational strategies to prevent misbehavior. Because student misconduct is the property of individuals, this preoccupation leads to models of management that are largely individualistic rather than group dynamics. Understanding the components that make up CM, Oliver *et al*. (2011) asserted that CM procedures are considered universal because they are implemented with the entire class rather than solely on individual students who may need additional support to reduce problematic behaviors.

Thus, beyond the commonly held view of establishing and enforcing classroom rules, routines, and procedures, Oliver *et al*. (2011) view CM as "A collection of classroom procedures implemented by teachers in classroom settings with all students to support pro-social behavior and prevent and reduce inappropriate behavior" (p.7-8). Brophy (1982) turned to the psychology of effective CM in that he distinguished between CM and behavior management, and he posited that successful CM "involves not merely responding effectively when problems occur but preventing problems from occurring very frequently at all. In turn, this prevention is accomplished primarily by good planning, curriculum pacing, and instruction that keeps students profitably engaged in appropriate academic activities" (p. 6). Teachers who implement the elements of CM, as defined by Oliver *et al*. (2011), to create a conducive learning

environment are more likely to produce positive student outcomes compared to those who adopt a disciplinary approach (Korpershoek *et al*., 2022).

Consistent with these arguments, the present study defines CM as the actions of preservice teachers to implement measures to address and correct disruptive behaviors as they occur (reactive approaches), establishing preventive strategies to minimize the occurrence of such behaviors (proactive approaches), and improving prosocial behavior by fostering a positive and respectful classroom atmosphere, which is established by enacting clear rules and procedures to provide structure and consistency, thereby supporting students' academic and social development.

1.1. Conceptualization of Teachers' Sense of Efficacy and Inclusive Classroom Management Self-Efficacy

Self-efficacy has often been conceptualized in copious literature grounded in Bandura's (1997) social cognitive theory as people's judgments of their capability to "organize and execute the courses of action required to produce given attainments" (p.3). Bandura asserted that the way people think and feel about their capabilities significantly influences their behavioral choices, the level of effort they exert, and their perseverance when encountering obstacles when pursuing and attaining their ambitious goals. The teachers' sense of efficacy (TSE), derived from Bandura's model, is a well-documented concept within the context of teacher education to IE, and it refers to teachers' beliefs about their capabilities to successfully undertake and execute teaching tasks to promote student engagement and learning (Tschannen-Moran *et al*., 1998; Tschannen-Moran & Woolfolk Hoy, 2001). While TSE is not only related to achieving desired outcomes of student learning and engagement, having efficacious beliefs about teaching is a requirement for teachers for their accomplishment and spurring actions (Armor *et al*., 1976; Tschannen-Moran *et al*., 1998).

The study of TSE as a motivational construct, reflecting teachers' competence beliefs regarding future endeavours, has maintained momentum since the 1970s. A plethora of studies have begun to investigate the critical role that TSE plays in the effective implementation of CM, which boosts student achievement (Guo *et al*., 2012; Hoy & Woolfolk, 1993; Tschannen-Moran *et al*., 1998). Initially, research on TSE focused on teachers' overall beliefs in their abilities to influence student outcomes, including their instructional practices, CM, and student engagement. As research delved deeper into the various domains of TSE, including instructional strategies and CM techniques, a conceptually distinct construct emerged: Classroom Management Self-Efficacy (CMSE) (Brouwers & Tomic, 2001). CMSE narrows the focus of TSE to specifically address the confidence teachers have in their skills and competencies needed to effectively manage classroom dynamics, maintain discipline, and establish routines to create a conducive learning environment (O'Neill, 2015). This evolution recognizes that while teachers may have a strong overall sense of efficacy, their confidence in managing classroom behaviors and interactions may vary. By incorporating this evolutionary perspective, the present study operationalized the concept of ICMSE by integrating the domains of CM and CMSE. As such, it refers to the beliefs that preservice teachers hold of their capabilities to effectively manage an inclusive classroom environment, ensuring that the diverse needs of all students are adequately met for their social and academic achievement and well-being.

The construct of ICMSE is significant to investigate as it directly relates to teachers' overall efforts in implementing effective CM strategies. It is reported that teachers with a positive sense of ICMSE are more likely to be motivated to perform management activities effectively in inclusive classrooms (Tschannen-Moran & Woolfolk Hoy, 2001). Ashton *et al.* (1983) posit that highly efficacious teachers are less likely to be affected by student misconduct and less likely to use harsh control strategies. Conversely, low efficacious teachers feel angered or threatened by student misbehavior and exhibit a more defensive than a humanistic approach to

CM. They further found that teachers with high self-efficacy are more likely to maintain "high academic standards, concentrate on academic instruction, monitor students' on-task behavior, and work to build friendly, non-threatening relationships with their low-achieving students than were teachers with low efficacy attitudes" (p. 23).

1.2. Assessing CMSE: Tracing its Roots

The construct of TSE is reported to have emerged in the 1970s through two distinct ideologies, Rotter's (1966) social learning theory, and Albert Bandura's (1997) social cognitive theory. Rotter's attribution theory concerns individuals' perceptions of their influence on the world around them. Rotter named this a "locus of control" and defined it as individuals' beliefs of controlling events. This control, whether internally within themselves or externally with others or situations, determines the consequences of their behavior. A second strand of research stems from the work of Bandura, who suggests that people's self-efficacy has two dimensions: Efficacy expectancy (belief about one's capabilities to perform a particular action) and outcome expectancy (one's belief about the likelihood of the consequences of that action).

The first notable study on TSE was conducted by the RAND Corporation, grounded on Rotter's social learning theory (Armor *et al.*, 1976). As part of a broader investigation into teacher characteristics, two 5-point Likert-type items were developed to assess teachers' beliefs about their influence on student outcomes. The items were: "When it comes right down to it, a teacher really can't do much – most of a student's motivation and performance depends on his or her home environment", and "If I try really hard, I can get through to even the most difficult or unmotivated students". Despite the modest scope of this research, RAND's inclusion of these items marked a pivotal moment, sparking broader interest and laying the foundation for the conceptual development of TSE.

The 1980s marked a pivotal shift in the development of multi-item instruments to measure TSE empirically. These Teacher Efficacy Scales (TES) expanded upon the original RAND items and integrated Bandura's (1977) social cognitive theory (Hoy & Woolfolk, 1993). These instruments typically adopted a two-factor structure that parallels Bandura's distinction between outcome expectations and efficacy expectations. The first factor reflects beliefs about the likely consequences of teaching behavior, closely aligned with the first RAND item, while the second factor pertains to personal beliefs in one's capabilities to execute effective teaching strategies, as reflected in the second RAND item. Some items in these scales, such as "If a student in my class becomes disruptive or noisy, I feel assured that I know some techniques to redirect him quickly," specifically capture the domain of CMSE.

These TES have become the predominant instruments for assessing TSE in an increasingly large number of studies undertaken in the subsequent years. In later years, the research by Hoy and Woolfolk (1993) provided additional evidence in support of the two-independent factorial structure of the TES, where the authors claimed that Factor 1 reflected teachers' general beliefs about the power of teaching rather than an outcome expectation of approaching difficult children, as asserted by Bandura (1997), Factor 2 represents teachers' sense of efficacy (Hoy & Woolfolk, 1993). Thus, the TES was adapted by Woolfolk and Hoy by labelling the first dimension as General Teaching Efficacy (GTE), reflecting beliefs about the general effectiveness of teaching, and the second dimension as Personal Teaching Efficacy (PTE), reflecting beliefs about one's own teaching abilities. A Woolfolk and Hoy version of the 22-item, two-factor structure of 6-point Likert-type TES (16 from Gibson and Dembo, four from their own and two from RAND) with the response categories ranging from "Strongly Agree" to "Strongly Disagree" has been used increasingly in their studies and other studies undertaken in subsequent years to measure these two TSE dimensions of preservice and experienced teachers (Hoy & Woolfolk, 1990, 1993).

Given the importance of CM, several studies emerged in the 1990s that developed instruments incorporating CMSE, either as subscales within the TES or as standalone measures. For

instance, Emmer and Hickman's TES added a subscale measuring CMSE to the original Gibson and Dembo's scale (Brouwers & Tomic, 2000). Of the 36 items on the TES, 14 represent teacher efficacy in CM and discipline construct and the sample items are "I find it easy to make my expectations clear to students", "I know what routines are needed to keep activities running efficiently" and "There are some students who won't behave no matter what I do". The CMSE subscale has been shown to have significant discriminant validity, distinct from the GTE and PTE subscales (O'Neill & Stephenson, 2011). Brouwers and Tomic (2000) used a Dutch-adapted version of the 14-item, 6-point Likert-type Self-Efficacy scale for Classroom Management and Discipline, originally developed by Emmer and Hickman, to examine the relation of perceived self-efficacy in CM and the three burnout dimensions among secondary (vocational) school teachers. In the following year, Brouwers and Tomic (2001) investigated the factorial validity of the 24-item, 6-point Likert-type Teacher Interpersonal Self-Efficacy Scale (TISES) using a sample of 832 in-service secondary teachers. This three-factor scale includes 14 items, adapted from Emmer and Hickman, representing perceived self-efficacy in CM subscale.

After reviewing several studies that employed instruments to measure TSE, Tschannen-Moran and Woolfolk Hoy (2001) proposed a promising new measure. They used two versions of the Teacher Sense of Efficacy Scale (TSES; formerly known as the Ohio State Teacher Efficacy Scale), the 24-item long form with three 8-item subscales and the 12-item short form. The subscales are the Efficacy for Instructional Strategies, Efficacy for Classroom Management, and Efficacy for Student Engagement, constructed on a 9-point Likert-type scale with anchored at "1 – Nothing", "3 - Very Little", "5 - Some Influence", "7 - Quite A Bit", and "9 - A Great Deal". The scale contains items in question form, and sample items in the Efficacy for CM subscale are "How much can you do to control disruptive behavior in the classroom?", "How much can you do to get children to follow classroom rules?", "How much can you do to calm a student who is disruptive or noisy?" and "How well can you establish a classroom management system with each group of students?"

The TSES has significantly influenced research on TSE for over a decade until the introduction of the 18-item, three-factor Teacher Efficacy for Inclusive Practices (TEIP) scale by Sharma *et al*. (2011). The TEIP scale is widely used in studies that address the unique challenges of inclusive classrooms, focusing on three TSE domains: Efficacy to use Inclusive Instruction, Efficacy in Collaboration, and Efficacy in Managing Behavior. Unlike the general focus of the TSES, the targeted approach of the TEIP makes it more relevant for assessing TSE in inclusive settings. For example, the behavior management subscale includes items, such as "I am confident when dealing with students who are physically aggressive" and "I am confident in my ability to prevent disruptive behavior in the classroom before it occurs", rated on a 6-point Likert scale from strongly disagree to strongly agree.

1.3. The Context and Research Gap

Regular class teachers in inclusive settings are reported to experience inappropriate behavior more often from students with diverse psychological, educational, and emotional needs than those working in non-inclusive settings (Milner & Tenore, 2010). This situation can be exacerbated when the class includes students from different cultural and social backgrounds (Aas *et al*., 2024; Main & Hammond, 2008; Sokal *et al*., 2013). The education system in Sri Lanka is no exception to this condition. As it gradually progresses toward achieving its full inclusion goals, most schools have begun to practice inclusion policies aimed at offering quality education for all while respecting the diversity of students and eliminating all forms of discrimination. This has created a learning environment more challenging for preservice teachers to deal with student disruptive behavior, which has been identified as one of the major causes for the increased levels of stress and burnout, as reported in local studies, for many teachers within and beyond the Sri Lankan education system (De Silva *et al*., 2015; Wickramasinghe *et al,* 2022; Yogaranee, 2024).

Although an exhaustive literature search resulted in several well-established instruments developed internationally as subscales to measure preservice teachers' self-efficacy in inclusive CM, the majority focused on the constructs of maintaining order and control, student socialization, and devising and implementing rules, expectations, routines, and procedures (Betoret, 2009; Brouwers & Tomic, 2001; Main & Hammond, 2008; O'Neill & Stephenson, 2011; Skaalvik & Skaalvik, 2007; Tschannen-Moran & Woolfolk Hoy, 2001). Most importantly, strong arguments about the items that referred to the conceptualization of inclusive/CMSE were lacking in much of the literature reviewed, as many items did not truly reflect this concept (O'Neill & Stephenson, 2011).

While research on ICMSE has advanced in recent years, there remains a significant gap in the availability of culturally and linguistically appropriate instruments to assess this construct, particularly in Tamil, one of the native languages of Sri Lanka. Developing a new scale specifically designed for preservice teachers, grounded in the cultural and linguistic nuances of their native language, is essential for accurately capturing this construct to manage diverse classroom environments effectively. Therefore, to bridge this gap, this study intends to develop a psychometrically sound instrument named "Tamil Inclusive Classroom Management Self-Efficacy (TICMSE) scale (*Thamizh Ulladakkiya Vakupparai Muhamaithuva Suya-Seyalthiran Alaveedu*) to measure the construct of ICMSE objectively. The following objectives guided this study: (1) to examine whether all developed items in the TICMSE scale truly measure the construct of ICMSE; (2) to assess that the TICMSE scale is multidimensional and (3) to establish its psychometric properties. More specifically, this study examines whether the manifested variables of the developed TICMSE scale are parsimonious in measuring the construct of ICMSE, not assessing the unrelated domains (Tucker & MacCallum, 1997). Addressing this gap also facilitates targeted interventions and training programs to support preservice teachers in preparing them for inclusive teaching.

2. METHOD

This methodological study is part of a larger experimental intervention study titled 'Effect of inclusive education course taught by flipped learning approach on pre-service teachers' self-efficacy'. Ethical approval for the main study was obtained (GS/ERC/2021/040) in December 2021 from the Faculty of Graduate Studies, University of Colombo.

2.1. Participants

The study's target population consisted of all preservice teachers enrolled in initial teacher education programs offered by National Colleges of Education (NCOEs) and state Universities in Sri Lanka. The accessible population, to which the findings can be generalized, consisted of 81% (629) preservice teachers from two NCOEs and 19% (144) from a state university. The TICMSE scale was administered online to a convenience sample of 240 preservice teachers, with 180 participants from NCOEs and 60 from the university. Two hundred and ten surveys were returned, yielding a response rate of 87.5%. Five responses were discarded due to the presence of outliers. Among the remaining valid responses, 97.5% of participants were female. Since male participants were underrepresented, their responses were excluded from the analysis, resulting in a final sample of 200 participants, including 156 from NCOEs and 44 from the university.

Various rules of thumb have been proposed in the literature to determine the minimum sample size required for Factor Analysis (FA). For example, Bentler and Chou (1987) recommended a ratio of N: p = 5:1, where N is the number of respondents, and p is the number of free parameters to be estimated, provided that all multivariate assumptions (such as normality and the absence of multicollinearity) are met. Hair *et al*. (2010) suggest that Structural Equation Modelling (SEM) typically requires a minimum sample size of 150 to 200 respondents for models of moderate complexity. However, the required sample size may vary depending on the complexity of the model, including the number of parameters and the quality and distribution

of the data. In addition, Hair *et al*. (2019) recommend 5 to 10 respondents per variable for Exploratory Factor Analysis (EFA).

This study finds a minimum sample size of 200 adequate for the model, involving 40 parameters (30-factor loadings, four-factor variances, and six covariances between the four factors in the final model), despite 250 participants that would be required to accommodate the 50 parameters (40 items, four-factor variances, and six covariances in the draft item scale). It is justifiable to conclude that if commonalities exceed the minimum threshold of 0.40, a sample size of 200 participants can serve as a practical lower limit for achieving reliable outcomes, particularly for models involving multiple latent variables (Costello & Osborne, 2005; Hair *et al*., 2019; Kline, 2016).

Before administering the survey, participants were informed about the confidentiality of the information shared, the potential risks and benefits of the study, and their rights to participate voluntarily or withdraw at any time.

2.2. Procedures Involved in the TICMSE Scale Development

2.2.1. *Item generation*

The process involved in the TICMSE scale development began with an extensive qualitative literature search to identify previously published instruments that captured the CMSE construct. During the search, it was confirmed that published instruments were developed grounded on Bandura's social cognitive theory (Tschannen-Moran & Woolfolk Hoy, 1998). In considering the conceptualization of self-efficacy, the literature consistently showed misconceptions and anomalies in constructing the scale items (Guskey & Passaro, 1994). Many instruments have truly measured the concept of opinion, views, perception, and other conceptions of self, such as self-concept, self-esteem, and self-worth, without bearing the conceptual underpinnings of Bandura (1997). As a result, researchers have questioned the validity and reliability of the instruments (Tschannen-Moran & Woolfolk Hoy, 2001). In the present study, careful consideration was given to phrasing the items representing teachers' sense of efficacy in CM and not any other teacher attributes (Guskey & Passaro, 1994). Thus, items were phrased with the referends "I can" and "I am able to do" as a statement of abilities reflecting efficacy as a conviction in performance (Slater & Main, 2020).

In addition, this study asserts that the items must encompass the component of efficacy expectations, as proclaimed by Bandura (1977). He notes that efficacy and outcome expectations are interrelated and function independently, but can be differentiated, where efficacy expectations precede and help form outcome expectations. If teachers believe, for example, that they can execute particular courses of action, they are more likely to initiate and perform the necessary actions at the expected level of competence to bring about desired results. On the other hand, "if they do not believe they can perform the necessary actions, they will not initiate the relevant behaviors or, if they do, they will not persist in those behaviors" (Guskey & Passaro, 1994, p. 4). A general template for the question reflecting the dimension of efficacy expectations is "Do I have the ability to organize and execute the actions necessary to accomplish a specific task at a desired level?" and for outcome expectations, "If I accomplish the task at that level, what are the likely consequences?" (See for example, Tschannen-Moran & Woolfolk Hoy, 1998, p. 210).

The next step in constructing the TICMSE scale was to select an appropriate format. The reviewed literature presented several formats, including forced-choice, vignettes, and rating scales. Guskey's (1986) alternative-weighting procedure for the Responsibility for Student Achievement (RSA) scale involved participants allocating 100 percent points between two alternative options for items that tapped into the RSA construct. Nevertheless, due to widespread use, ease of administration, and interpretation, this study identified the Likert-type rating scale as appropriate for constructing the scale (Cohen *et al*., 2007).

The TICMSE scale was developed using procedures suggested by DeVellis (2017) and Boateng *et al*. (2018). Both practices were highly influential in facilitating the development of new, valid, and reliable scales and improving those already established in various disciplines, including the health, social, and behavioral sciences. The TICMSE scale development began with an item generation process by creating an item pool consisting of author-generated items and items drawn from the existing literature. The literature shows inconsistent results for the item format of the scale. Some scales were in statement form, while others were in question form. The most widely recommended and used format is the Likert-type, self-report rating scale, usually in statement form. Hence, the present study selected the statement form and finalized the scale with 40 items. Although alternating wording of items has some advantages, including elimination of acquiescent and extreme response biases, the study adopts a positive wording format because it reduces misinterpretation of items on the part of the respondents and miscoding of items on the part of a researcher.

The third step was to select the anchoring points and labels appropriate for the item types. The results in the literature have been inconsistent in utilizing the response categories for Likert-type rating scales (Spector, 1992). Many have used agreement-type labels anchored at "strongly agree" to "strongly disagree" with five, six, seven, and nine rating points. Others have focused on a 9-point scale anchored at "Nothing", "Very little", "Some influence", "Quite a bit", and "A great deal" (as in Bandura, n.d.). The selection of an even number scaling system with no midpoint has different arguments among researchers, where some argued that there is a central tendency and a regression to the mean issues when having a midpoint, while others believed that it gives freedom to "those who wish to sit on the fence and choose a midpoint" (Cohen *et al*., 2007, p. 327). Since there is no rule to strictly follow the number of response categories in the Likert-type rating scale, the present study adopted a six-point scale, devoid of a mid-value. The scale was anchored at "*Strongly Agree*", "*Agree*", "*Somewhat Agree*", "*Somewhat Disagree*", "*Disagree*", and "*Strongly Disagree*". These categories were deemed appropriate for measuring the attitudinal component of self-efficacy. Ratings were assigned on a scale of one to six, with the highest rating corresponding to "*Strongly Agree*."

2.2.2. *Pretesting the first draft of the 40-item TICMSE scale*

After establishing the format of the scale and the items, the subsequent step involved generating items from the item pool. This process resulted in 40 items that effectively represent the ICMSE construct. As conceptualized in this study, four factors were predefined, with ten items allocated to each dimension. The pre-final version of the TICMSE scale was then piloted with ten preservice teachers, not representing the target audience. After modifications to the pilot results, including rewording items with ambiguities, discarding confusing items, and correcting instructions to ensure they are easily understandable to the target audience.

Subsequently, the scale underwent pretesting with a small subsample of preservice teachers to assess the articulation and clarity of words and phrases, the representativeness and comprehensiveness of items, and the appropriateness of response categories. Pretesting is crucial to ensure that each respondent interprets the items as the researcher and other respondents (DeVellis, 2017). The analysis of the pre-test facilitated the reformulation of certain items. Among the 40 items, two exhibited ceiling effects with many participants selecting the extreme response options. Two items demonstrated floor effects, indicating a significant proportion of participants achieved the lowest scores. As a result, these four items were removed from the scale to prevent a loss of variability in the data. The remaining 36 items were retained in the first draft of the scale for content validation.

2.2.3. *Establishing the content validity for the 36-item TICMSE scale*

Quantifying the content validity of a scale is a prerequisite for establishing other qualities of instruments, including construct validity and reliability (Lamm *et al*., 2020). Content validation of a psychological instrument refers to the process that intends to ensure an instrument measures

the content area that it purports to measure (Ayre & Scally, 2014). Polit and Beck (2006) and DeVellis (2017) define content validity as sampling adequacy, which refers to the extent a sample of items in a psychological instrument accurately reflects the construct being assessed. The content of the scale can be quantified using the Content Validity Ratio (CVR) and the Content Validity Index (CVI) (Lawshe, 1975). In the literature, the most widely used and reported approach to content validity is the CVI due to its simple calculation and interpretation and its agreement of relevance instead of agreement per se, on top of providing both item and scale validation measures (Polit *et al*., 2007). However, a common criticism of the use of CVI reported in the literature is that it fails to adjust for chance agreement (Polit & Beck, 2006). Inter-rater reliability can be tested using Fleiss' kappa, a statistical measure designed to evaluate agreement among multiple raters. The interpretation of kappa values followed the widely cited benchmarks proposed by Landis and Koch (1977), where values less than 0.00 indicate poor agreement; 0.00–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, almost perfect agreement.

In this study, a content evaluation panel comprising six subject experts, three from the field of IE and three from special education, was established to assess the content validity of the TICMSE scale. Each expert evaluated the relevance of every item using a 4-point Likert-type scale, where 4 indicated “highly relevant,” 3 “quite relevant,” 2 “somewhat relevant,” and 1 “not at all relevant.” The item-level content validity index (I-CVI) was then calculated by dividing the number of experts who rated the item as either 3 or 4 by the total number of experts. Following the guidelines by Polit *et al.* (2007), an I-CVI value of ≥ .78 was considered indicative of excellent content validity.

Accordingly, items with I-CVI values between .50 and .77 were revised. Six items found to have the least relevance, with I-CVI values of .50 and below, were excluded, resulting in 30 items for which content validity was established. For reference, the Tamil version of the TICMSE scale along with its English translation is available in Supplementary.The Scale-level Content Validity Index for Average (S-CVI/Ave) refers to the average of the item-level CVIs (I-CVIs) across all items. This index was computed using the average method, and the results showed a value of .96, which is above the .90 threshold, indicating that the scale has excellent content validity (Polit & Beck, 2006). The Scale-level Content Validity Index for Universal Agreement (S-CVI/UA) is the proportion of items that received a rating of 3 or 4 by all experts. This index was also computed. The results showed a value of .77, which is ≥ .70 threshold, indicating that, on average, about 77% of the items on the scale are considered valid by all the experts. To assess the effect of chance agreement among the raters, an inter-rater agreement analysis was conducted to determine the reliability of the ratings of the six raters. For this purpose, Fleiss’s kappa, a measure of agreement between more than two dependent categorical samples, was computed. The k value of .61 falls within the range of .61 - .80 benchmarks, indicating a substantial agreement among raters (Landis & Koch, 1977).

2.2.4. *Establishing the factorial structure of the 30-item TICMSE scale*

Data analysis was conducted in two phases. In the first phase, an EFA was performed using Jamovi (version 2.3) software to determine the factorial structure of the 30-item TICMSE scale for which the data were collected from 200 preservice teachers. EFA, a data reduction technique, systematically examines whether many observed variables are linearly related to a small set of variables, preferably uncorrelated, or to groups of unobservable (latent) factors. This grouping is based on the intercorrelations among a set of observed variables, determined by commonality, which represents the proportion of variance in each observed variable that is accounted for by the common factors (Hair *et al*., 2019). Items with communality values ≤ .50 are typically dropped from the scale (Field, 2013). A common application of EFA is to develop a scale to objectively measure an unobservable psychological construct that cannot be directly measured (Cheung *et al.,* 2024). Latent variables are then used to parsimoniously explain the covariation observed among a set of manifested variables (Watkins, 2018).

Initially, the factorability of the correlation matrix was subjectively assessed, following the guidelines by Tabachnick and Fidell, (2013). According to their recommendation, if no correlations ≥ .30 are found, factor analysis of the matrix is considered meaningless. Hair *et al*. (2010) suggest that a sizable number of inter-correlation coefficients (≥ .30) is adequate for EFA. Multicollinearity, indicated by coefficients ≥ .80, and singularity of the observed variables were also assessed (Field, 2013). The factorability of the correlation matrix can be objectively evaluated using Bartlett's Test of Sphericity and the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO-MSA) test (Taherdoost *et al*., 2014). Bartlett's Test of Sphericity produces a chi-square value, indicates the statistically significant relation of the correlation matrix of original measured variables to the identity matrix, which contains 'ones' on the principal diagonal and 'zeros' on the off-diagonals, where each variable is uncorrelated with other variables while retaining a perfect self-correlation.

Since the Bartlett test is sensitive to large sample sizes, it may produce a significant statistic even with fairly trivial correlations within the matrix. Therefore, it is recommended that this test be supplemented with the KMO-MSA statistic. According to Kaiser's (1974) description, the KMO-MSA refers to the "ratio of correlations and partial correlations that reflects the extent to which correlations are a function of the variance shared across all variables rather than the variance shared by particular pairs of variables" should not be ≤ .50, suggests an inappropriate correlation matrix for factoring (Watkins, 2018, p. 226). Furthermore, the appropriateness of the sample data for EFA was determined by analysing the MSA of individual items, which should be ≥ .50, and the initial estimation of commonalities in the potential factor solution (Hair *et al.,* 2019).

After establishing the suitability of the correlation matrix for FA, the next step was to determine the number of factors to retain in the final scale. This was accomplished through various measures, including examining eigenvalues, the scree plot, parallel analysis, and factor loadings with a minimum threshold ≥ .40 (Costello & Osborne, 2005; Field, 2013; Hair *et al*., 2019). The 30-item TICMSE scale was factor analysed using the Principal Axis Factoring (PAF) method as the extraction method, Direct Oblimin (DO) as the oblique rotation method, and parallel analysis to determine the factor solution. The reason for selecting PA over Maximum Likelihood (ML) is that the former outperforms when dealing with weak factor loadings (≤ .40) and small sample sizes (≤ 300), as well as when the multivariate normality assumption is not met (Briggs & MacCallum, 2003; Curran *et al*., 1996; MacCallum *et al.,* 2001). An oblique rotation method allows the factors to correlate to improve the intercorrelations between the items within the factors. By re-running EFA after deleting the items with poor factor loadings, the final factor solution for the scale was produced.

2.2.5. *Assessing the fit of the 27-item TICMSE scale*

In the second phase, a Confirmatory Factor Analysis (CFA) was conducted using IBM SPSS AMOS (*v*.23) software with ML estimation to confirm whether the measured variables accurately represent the latent variables and establish the scale's psychometric properties. Initially, standardized lambda was estimated to determine the items that adequately loaded onto the respective latent variables. Stepwise item deletion was then performed for items showing standardized factor loadings ≤ .40 (Kline, 2016). Finally, modification indices were estimated to refine the measurement model and determine whether the hypothesized model fits the observed data. The chi-square statistic (χ^2), Goodness of Fit (GFI) and absolute fit indices (Root Mean Square Error of Approximation, RMSEA and Standardized Root Mean Square Residual, SRMR) and the Adjusted Goodness of Fit (AGFI), the Comparative Fit Index (CFI), the Tucker Lewis Index (TLI; also called as Non-Normed Fit Index -NNFI) as incremental fit indices and χ^2/df as parsimonious fit index gaged model fit. The CMIN/df was considered to confirm the overall model fit if the value was ≤ 2 for good fit and ≤ 3 for acceptable fit. The threshold values were set at ≤ .06 for RMSEA (Hu & Bentler, 1999), ≥ .90 for GFI and AGFI (Byrne, 2010), ≥ .95 for CFI, TLI (Hu & Bentler, 1999), and ≤ .08 for SRMR (Hu & Bentler, 1999).

3. RESULTS

3.1. Results of EFA

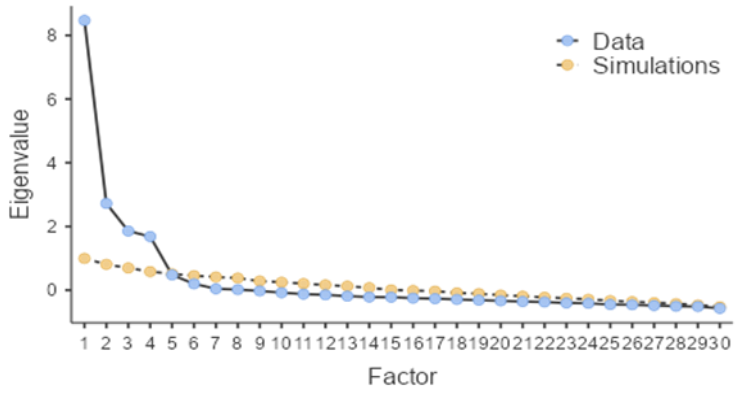
Before performing EFA, the data were examined for multivariate assumptions, including outliers, normality, and linearity. Multivariate outliers were assessed using a Mahalanobis distance test, which revealed that the 30-item dataset had no multivariate outliers, as the *p*-values for each item were ≥ .001 (Tabachnick & Fidell, 2013). The assumption of multivariate normality was met, with skewness and kurtosis statistics falling within an adequate range (between -1 and +1) (Hair *et al*., 2017). This finding was further supported by the Shapiro-Wilk test, which confirmed the normality of the data with a W statistic of .986 ($p \geq .05$), which was greater than the critical value of .932 ($p = .05$) and .818 ($p = .05$), indicating that the overall data were normally distributed (Shapiro & Wilk, 1965). Scatter plots were examined to assess the assumption of linearity by visualizing the relationships between factors and identifying deviations from linear relationships.

Examination of the correlation matrix revealed no evidence of multicollinearity or singularity issues, as no items had bivariate coefficients (r) ≥ .80 or equal to one. However, a few items had an r value < .30, contributing less than 10% of the variation to the latent construct. These underperforming items were supposed to be deleted from the scale. However, as it was an initial subjective qualitative analysis of the data, deletion was delayed until the factorability of the matrix. The multicollinearity assumption was statistically assessed using the Variance Inflation Factor (VIF) and tolerance values for each observed variable. The cutoff below .10 (or .20) for tolerance values and above 10 (or 5) for VIF indicates problematic multicollinearity (Kutner *et al*., 2004). The results showed that all predictor variables exhibited low VIF and high tolerance values close to one, indicating minimal multicollinearity issues.

The chi-square value produced by Bartlett's Test of Sphericity was statistically significant ($\chi^2_{(435)} = 3311, p < .001$), indicating that the sample correlation matrix was significantly different from the identity matrix, implying that the observed variables were significantly correlated to each other. Additionally, this study found that the overall KMO MSA was .893, and the individual variables' KMO MSA was ≥ .50, indicating that the correlation matrix was "meritorious" (Kaiser, 1974). All these statistics, along with the subjective evaluation conducted on the correlation matrix, suggested that the sample data on hand was factorable.

The EFA conducted using unrotated PCA yielded six factors, determined by Kaiser's eigenvalue ≥ 1 rule, which refers to the total variance accounted for by each factor. Collectively, these factors explain 66.21% of the variance. This is further supported by the scree plot (Figure 1), which helps to identify the optimal number of factors to retain in the scale. Upon visual inspection of this plot, a slight break was observed after the fourth factor, with another noticeable break at the fifth factor. Additionally, a parallel analysis (simulated data represented by a horizontal yellow dotted line) supported this four-factor solution of the scale. The EFA was conducted again using the PAF with DO rotation technique, focusing on the four factors identified earlier. The resultant factor matrix is presented in Table 1. Accordingly, all items loaded perfectly onto the respective factors, except for a few items, which demonstrated poor as well as cross-loadings with values ≤ .40. The visual inspection of the Scree Plot along with parallel analysis and Kaiser's eigenvalue ≥ 1 rule, suggested a final four-factor solution. This solution explained 52.7% of the total variance in the observed data. This model confirms that the data thoroughly represented the underlying four-factor structure of the construct, ICMSE ($\chi^2(321) = 475, p < .001$).

Figure 1. *The scree plot with parallel analysis.*



3.1.1. *Naming the dimensions*

As many teachers and teacher educators have conceived, the concept of CM does not embrace the idea of controlling student behavior in a mechanical, authoritative style to maintain discipline, compliance, and obedience in the classroom (Güner-Yıldız *et al*., 2022). For them, this term is typically linked with reactive management efforts taken by teachers, such as responding to misbehavior with rewards and punishments and maintaining order through silence. However, researchers have consistently questioned the efficacy of these responsive approaches to manage classrooms (Evertson & Poole, 2008; Korpershoek *et al*., 2016).

Numerous studies have consistently argued that CM should be conceptualized beyond traditional boundaries, as it is broader in scope than implied by terms such as discipline or control. It is a more complex and versatile effort to prevent and address student misbehavior while establishing rules, procedures, and regulations to create a positive learning environment that fosters students' academic and social-emotional development. Thus, CM is considered a key condition for student learning and achievement. To this end, preventive and remedial actions are critical elements, although the former is generally considered more effective than the latter (Korpershoek *et al*., 2022; Main & Hammond, 2008; Marzano *et al*., 2003). According to Webster-Stratton *et al*. (2011), students' positive behaviors should get more attention from teachers than negative behaviors.

Therefore, in addition to reactive actions, the present study incorporates elements of proactive (e.g., encouragement and praise to ensure that students are engaged) and regulating management efforts, as well as actions aimed at improving student prosocial behavior, which together constitute the concept of CMSE to ensure meaningful learning. Accordingly, four dimensions of inclusive classroom management and their corresponding self-efficacy constructs for preservice teachers were defined. These dimensions were named (1) self-efficacy for reactive actions (SERA), (2) self-efficacy for proactive actions (SEPA), (3) self-efficacy for improving prosocial behavior (SEPB), and (4) self-efficacy for enforcing classroom rules and procedures (SECRP).

SERA refers to teachers' confidence in their abilities to implement various reactive disciplinary strategies, including, warnings, reprimands, or punishments for disruptive students. These strategies are reported to be used most frequently by teachers, although their efficacy in effectively changing student behavior is questioned. Studies have asserted that teachers are more likely to resort to presumably less effective reactive strategies due to inadequate knowledge, lack of confidence, or uncertainty about the efficacy of preventive measures (Peters, 2012; Woodcock & Reupert, 2012; Yogaranee, 2024).

The literature consistently shows the critical role that preventive actions of teachers play in over reactive strategies for effective CM (Korpershoek *et al*., 2022). In this case, SEPA is included as a factor of ICMSE, referring to the beliefs teachers have about their capabilities to prevent student disruptive behavior. Proactive strategies typically aim to anticipate and mitigate

potential disruptions, such as misbehavior, noncompliance, or disengagement, that may interfere with the instructional process. According to Goss *et al*. (2017), high expectations, strong teacher-student relationships, clarity and structure in instruction, and active learning are considered the most effective strategies used by teachers to prevent student misbehavior by keeping their focus on learning.

Teachers' proactive actions are more feasible in a student-centred learning culture, as they emphasize the integration of teaching and discipline. An interesting, stimulating, and active learning environment reduces the risk of disruptive behavior. Additionally, in these contexts, students are more likely to understand what behavior is expected in the classroom, provided that teachers are thoughtful and proactive in their actions. This can include actions taken before students arrive and interactive activities that teachers plan once students arrive, intended to minimize disruptions, distractions, and interference with the smooth conduct of teaching and learning (Evertson & Poole, 2008). On the other hand, the SEPB dimension included in the ICMSE concept was derived from the premise that preservice teachers who exhibit high levels of self-efficacy in CM by building positive relationships with students and fostering conducive learning environments serve as powerful role models for their students. When students witness their teachers' assertive and appealing behaviors, they are more likely to emulate these traits in their behavior. Fostering prosocial behavior among students is one of the powerful socialization processes that teachers adopt to curb problematic behaviors.

Research has consistently shown that the student socialization process, by fostering prosocial behavior, is not mutually exclusive to other preventive and reactive measures; it can be incorporated into both or either of these approaches (Korpershoek *et al*., 2022). The present study, however, regards this dimension as a distinct entity and does not categorize it within the scope of the preventive approach. This is because prosocial behaviors are desirable and contribute to a positive atmosphere, rather than posing a challenge or threat to the learning environment. With that said, this self-efficacy dimension encompasses teachers' abilities to promote students' good conduct rather than reducing or preventing misconduct.

Classroom rules should not be imposed but negotiated with students and implemented with mutual consent by teachers and students. Several studies have reported teachers' regulating actions to develop and enact classroom rules and procedures as preventive measures (Marzano *et al*., 2003). However, proactive approaches emphasize prevention or minimizing disruptions while establishing rules and regulations focusing on setting clear classroom expectations and consequences for student behavior. Therefore, in the present study, SECRP as a distinct domain of the ICMSE refers to the beliefs teachers hold about their capabilities to enact classroom rules, routines, and procedures to accomplish their educational objectives by improving students' behavior and enhancing their academic performance. These rules constitute clear expectations and consequences for student behavior across various activities, including general in-class behavior, group work, seat work, and the use of learning materials and equipment from the beginning of the period until the end of the period or school day.

This categorization of the dimensions for the concept of ICMSE aligns with four of the six personal teacher action categories of O'Neill and Stephenson (2011). They are (1) "Establishing and maintaining rules, routines, procedures and expectation" (regulating inappropriate behavior by enacting rules and procedures as specified in the present study– SECRP), (2) "Facilitating student socialization and cooperation" (fostering prosocial behavior as specified in the present study - SEPB), (3) "Gaining and maintaining attention and monitoring engagement in tasks" (proactive actions as specified in the present study – SEPA) and (4) Maintaining order and control" (disciplinary/reactive actions as specified in the present study - SERA).

The sample item included in the SECRP subscale is "I can get students to follow classroom rules and procedures through consistent reinforcement and feedback."; the SEPB subscale, "I know how to develop positive relationships with students to make it easy for them to approach me."; the SEPA subscale, "I can teach students calming techniques and self-regulation

strategies to help them deal with negative emotions and strong impulses." and the SERA subscale, "I am confident that I can verbally reprimand and correct students for disruptive behavior rather than resorting to punishment." The items that refer to the broad abilities of teachers to the nonspecific actions of students, for example, "I am able to implement classroom management skills effectively," "I can control student misconduct," or "I can effectively implement behavior management strategies to handle student misbehavior" were not included in the scale as they encompass all the actions taken by teachers to managing student misbehavior. Moreover, the construct of ICMSE in this study mainly focuses on the personal self-efficacy of preservice teachers rather than a generalized one.

After identifying and labeling the four key dimensions of the developed scale, each item was systematically coded to reflect its corresponding dimension and sequence. Accordingly, Arabic numerals were assigned to the individual items within each subscale. For example, items under the SERA subscale were coded as SERA 1, SERA 2, etc. A similar coding system was followed for the other scale domains, SEPA, SEPB, and SECRP, to ensure clarity and ease of reference for data analysis and interpretation of results.

3.1.2. *The final factor solution for the TICMSE scale*

The final run of EFA, conducted after removing items SEPB4, SEPB8, and SERA4 due to weak and cross-loadings, yielded a 27-item, four-factor solution ($\chi^2(249) = 319$, $p < .001$). Furthermore, the model included an adequate number of items in each dimension, each having no fewer than three items (Brown, 2015; Guadagnoli & Velicer, 1988). The percentages of variance accounted for by SECRP, SEPA, SERA, and SEPB were 15.6%, 14.9%, 14.5%, and 11.4%, respectively. Together, these four latent variables explained the cumulative variance of 56.4%, which was deemed acceptable (DeVellis, 2017; Hair *et al*., 2019).

Table 1. *Initial factor solution for the 30-item TICMSE scale with four extracted factors.*

		Factors			
Items		1	2	3	4
SECRP6	I can seek support or resources when facing challenges in enforcing classroom rules.	**0.875**	0.002	0.019	-0.023
SECRP4	I can get students to follow classroom rules and procedures through consistent reinforcement and feedback.	**0.804**	-0.006	-0.008	-0.043
SECRP2	I can consistently enforce classroom rules and procedures.	**0.743**	0.033	0.058	-0.022
SECRP1	I can establish classroom rules related to learning tasks, such as completing assignments on time, coming to class prepared with necessary materials, and participating in group activities or discussions when asked at the beginning of the school academic year.	**0.735**	0.085	0.055	-0.039
SECRP5	I can effectively handle differences of opinion regarding classroom rules and procedures.	**0.702**	-0.005	0.065	0.03
SECRP7	I can adapt classroom rules and procedures to accommodate the diverse needs of my students.	**0.694**	0.066	-0.023	0.089
SECRP8	I can effectively develop rules related to respect, such as raising hands to speak, listening to others' speech treating fellow students with respect, and using appropriate language and the consequences of violating them.	**0.576**	-0.158	-0.09	0.091
SECRP3	I can act fairly and impartially in enforcing classroom rules.	**0.441**	0.003	-0.081	0.071
SEPB8	I can be a good role model by demonstrating respect, kindness, and self-control in my interactions with students.	0.376	-0.145	-0.06	0.263
SEPB4	I can effectively implement peer teaching to support the learning of slow learners in my class.	0.316	-0.006	-0.114	0.228
SERA5	I can de-escalate conflicts between students in my class by taking constructive resolutions.	-0.01	**0.872**	-0.044	0.005

SERA3	I can openly communicate with parents of students with severe behavioral problems, such as using or being under the influence of alcohol or drugs and seek their support in addressing such behaviors.	0.011	**0.826**	-0.011	-0.051
SERA1	I feel confident in my ability to promptly address unexpected disruptions by students to learning in the classroom, such as talking out of turn, making noise, frequently getting up and moving around the classroom without permission, or engaging in activities that distract others from learning.	-0.060	**0.805**	0.106	0.035
SERA6	I can seek support from school administrators to address persistent or severe behavioral problems of students in my class.	-0.030	**0.741**	0.004	0.067
SERA7	I am confident I can verbally reprimand and correct students for disruptive behavior rather than resorting to punishment.	0.157	**0.738**	-0.027	0.017
SERA2	I believe I can effectively handle the destructive behaviors of students such as vandalizing classroom properties, and intentionally damaging materials or equipment in the classroom.	-0.010	**0.733**	0.038	0.008
SERA4	I know how to deal with the disrespectful behaviors of students in my classroom, such as talking back to the teacher, arguing, or showing a lack of respect for classmates through rude comments, gestures, or body language.	0.188	**0.393**	-0.058	0.051
SEPA4	I can prevent disruptions by making students feel valued and respected by recognizing their responsible behaviors in the classroom.	0.009	-0.063	**0.835**	-0.048
SEPA5	I can create a supportive classroom environment that empowers students to take responsibility for their learning and behavior.	-0.02	0.031	**0.810**	-0.069
SEPA3	I can teach students calming techniques and self-regulation strategies to help them deal with negative emotions and strong impulses.	-0.03	-0.030	**0.771**	0.072
SEPA6	I believe that I can listen attentively to students' concerns and opinions.	0.016	0.068	**0.715**	-0.038
SEPA1	I can effectively teach in the classroom to prevent students from engaging in unmotivated behaviors such as daydreaming or not paying attention to the lesson.	-0.010	-0.013	**0.701**	0.135
SEPA2	I can anticipate potential issues or disruptive behaviors in the classroom and take appropriate steps promptly to handle them before they escalate.	0.052	0.076	**0.693**	0.049
SEPA7	I can create learning opportunities for students with special educational needs by using differentiated instruction to meet their various needs.	0.138	0.053	**0.605**	0.066
SEPB6	I know how to develop positive relationships with students, making it easy for them to approach me.	0.048	-0.01	-0.011	**0.757**
SEPB7	I can create opportunities for students to improve their selfless behaviors, such as sitting together or eating together with fellow students who feel excluded or isolated in class.	-0.003	-0.015	-0.025	**0.743**
SEPB1	I can improve students' sharing behaviors by encouraging them when they willingly share their belongings, resources, or time with fellow students in the class.	-0.055	0.094	-0.031	**0.695**
SEPB2	I believe I can develop students who value diversity through my effective inclusive classroom teaching.	-0.020	0.042	0.023	**0.673**
SEPB5	I can provide students with adequate opportunities to participate in group learning that fosters social consciousness rather than relying solely on individual learning in the class.	0.038	-0.002	0.088	**0.619**
SEPB3	I am confident that I can empathize with students who exhibit aggressive behavior rather than punishing them.	0.016	0.009	0.140	**0.612**

Note. 'Principal axis factoring' extraction method was used in combination with an 'Oblimin' rotation

The number of items retained in the 27-item TICMSE scale is as follows: The SECRP subscale comprised items SECRP1, SECRP2, SECRP3, SECRP4, SECRP5, SECRP 6, SECRP7, and SECRP8 (n = 8); the SEPB subscale comprised items SEPB1, SEPB2, SEPB3, SEPB5, SEPB6, and SEPB7 (n = 6); the SERA subscale comprised items SERA1, SERA2, SERA3, SERA5, SERA6, and SERA7 (n = 6); and the SEPA subscale comprised items SEPA1, SEPA2, SEPA3, SEPA4, SEPA5, SEPA6 and SEPA 7 (n = 7). The wording of the items in the scale was positive; therefore, the response categories for all items were coded on a scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Computing the correlations to demonstrate the linear association between each pair of variables within the individual factors revealed significant positive Pearson's r values, indicating positive correlations. The reliability analysis conducted to compute Cronbach's alpha yielded excellent estimates of .881, .902, .913, and .855 for the subscales SECRP, SEPA, SERA, and SEPB, respectively. These values exceeded the benchmark of .70, suggesting a strong internal consistency of the subscale scores (Hair *et al*., 2010). Analysis of inter-factor correlations for the scale indicated that the factors exhibited some moderate correlations with each other, with correlation coefficients ranging from .32 to .45. With this confirmation, the 27-item TICMSE scale was recommended for CFA to validate its four dimensionality and establish significant psychometric properties, including composite reliability, discriminant and convergent validity, as discussed in the following sections.

3.2. Results of CFA - Refinement of the 27-item TICMSE Scale

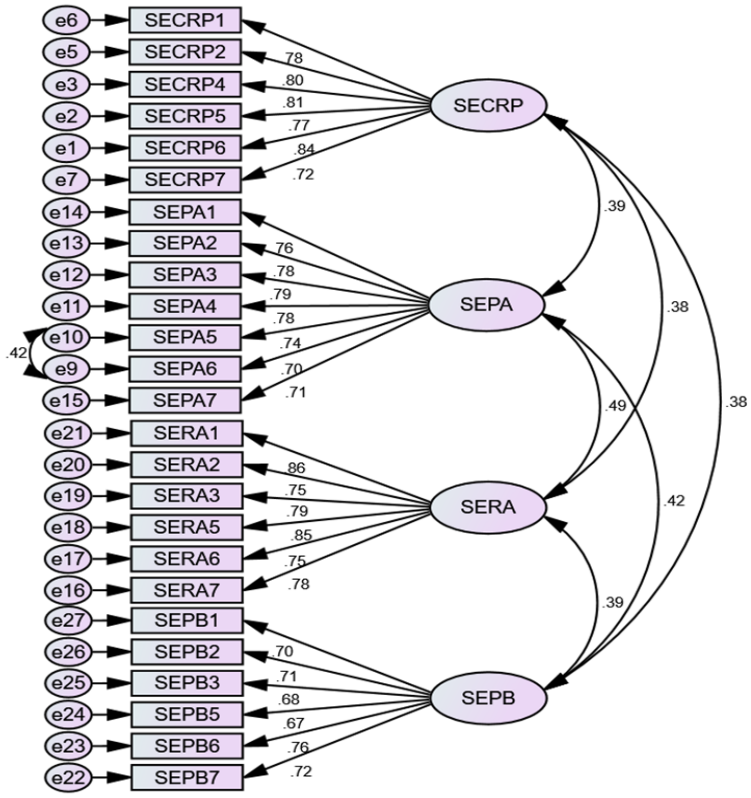
CFA was conducted to evaluate how well a set of 27 items reflects an a priori assumption of the four-factor structure of the TICMSE scale. The results indicated that the initial measurement model had a statistically significant chi-squared value ($\chi^2_{(318)} = 418.193$, $p < .000$), with model fit indices suggesting good fit overall, except for GFI and AGFI (see Table 2). Specifically, CMIN/*df* = 1.315, RMSEA = .040, CFI = .965, GFI = .867, TLI = .961 SRMR = .527, indicative of satisfactory fitness. Moreover, the standardized factor loadings for all indicators were significant, with loadings ≥ .70, except for SECRP3 and SECRP8. These two items had very low variances of 19% (.431) and 26% (.507), respectively, despite their significant loadings ($p < .000$). Additionally, SEPB1, SEPB3, and SEPB5 had a factor loading slightly below, close to the .70 threshold (ranging from .67 to .69). The decision was made to retain these items until further evaluation of the scale because they demonstrated significant loadings onto the SEPB dimension.

After stepwise removal of SECRP3 and SECRP8 from the model and covariating the error variance as shown in Figure 2, the final measurement model yielded estimates as presented in Table 2. All 25 items were adequately loaded onto the respective latent variables, with loadings above the threshold value of .70, except for SEBP3 and SEPB5, which had std. lambda values below but close to the .70 threshold (.68 and .67 respectively). Additionally, the model exhibited a significant chi-squared value ($\chi^2_{(268)} = 317.349$, $p = .021$), indicating that the model did not fit the data well. This suggests that the discrepancies between the observed and model-implied covariance matrices are statistically significant (Kline, 2016). However, it's important to note that the χ^2 test can be sensitive to sample size, particularly with large sample sizes (Bentler & Bonnet, 1980; Hooper *et al*., 2008). Therefore, alternative measures of model fit, such as Wheaton *et al*.'s (1977) relative/normed chi-square (CMIN/df) have been proposed. Since there is no agreed-upon threshold value for this statistic, it is recommended that ranges between as high as 5.0 and as low as 2.0 can be acceptable (Tabachnick & Fidell, 2013).

The model fit indices indicate a good fit, with a CMIN/df of 1.184 (≤ 2), a CFI of .982 (≥ .95), a TLI of .980 (≥ .95), an RMSEA of .030 (≤ .06), and an SRMR of .0478 (≤ .08) (Table 2). However, it is worth noting that both the GFI and AGFI indices provided inadequate fit estimates (.890 and .866, respectively), as they did not meet the threshold of ≥ .90 (Byrne, 2010). Further refinement of the measurement model may be necessary to improve these

indices. Nonetheless, it is important to recognize that while GFI and AGFI were widely used in the past, they are often supplemented by more robust and informative fit indices such as CFI, TLI, RMSEA, and SRMR (Hooper *et al*., 2008). Therefore, this study does not rely on GFI and AGFI when reporting the results of the CFA.

Figure 2. *Final measurement model after deletion of SECRP3 and SECRP8 and Covariate Error Variances.*



Note: Factor 1 - Self-efficacy for enforcing classroom rules and procedures (SECRP), contains items 1,2,4,5,6 and 7 (n=6); Factor 2 - Self-Efficacy to adopt Reactive Approaches to dealing with inappropriate behavior (SERA), consists of items 1,2,3,5,6, and 7 (n=6); Factor 3 - Self-Efficacy to adopt Proactive Approaches to dealing with inappropriate behavior (SEPA), contains items 1,2,3,4,5,6 and 7 (n=7); and Factor 4 - Self-Efficacy to improve Prosocial Behavior (SEPB), contains items 1,2,3,5,6, and 7 (n=6).

Table 2. *Fit indices, corresponding thresholds and interpretations.*

Fit indices	First MM Estimates	Final MM Estimates	Threshold values	Interpretation
χ^2	418.193	317.349	Insignificant at .05	Badness of fit
df	318	268	threshold (Hu & Bentler,	
p	< .000	< .05	1999)	
CMIN/DF	1.315	1.184	≤ 2 – absolute fit ≤ 3 – acceptable fit	Absolute fit
RMSEA	.040	.030	≤ .06 (Hu & Bentler, 1999)	Good fit
GFI	.867	.890	≥ .90 (Byrne, 2010)	Poor fit
AGFI	.841	.866	≥ .90 (Byrne, 2010)	Poor fit
SRMR	.0527	.0478	≤ .08 (Hu & Bentler, 1999)	Acceptable fit
CFI	.965	.982	≥ .95 (Hu & Bentler, 1999)	Good fit
TLI (NNFI)	.961	.980	≥ .95 (Hu & Bentler, 1999)	Good fit

3.3. Validation of the 25-item TICMSE Scale Scores

Once the hypothesized, a priori four-factor structure was confirmed using CFA to converge appropriately to produce the final dimensionality of the TICMSE scale, the next step was to evaluate its score reliability and validity. Among the many reliability and validity measures of psychological instruments available in the literature, construct reliability and construct validity measures are considered important when reporting a CFA measurement model using SEM (Hair *et al*., 2014). Reliability refers to the consistency of the scale scores, meaning that the extent to which the scale yields consistent results with its repeated administration on different occasions using the same sample (Hair *et al*., 2010). Construct reliability refers to the degree to which the indicator variables consistently measure the underlying latent construct. Cronbach's alpha and composite reliability (CR) are commonly used to assess the construct reliability. CR is computed for each factor using the equation $CR = (\sum\lambda_i)^2 / (\sum\lambda_i)^2 + \sum\theta_i$, interpreting as the sum of the standardized factor loading (λ_i) squared divided by the sum of standardized factor loading (λ_i) squared plus the sum of the error variance (θ_i) (Cheung *et al*., 2024). As a rule of thumb, a CR value ≥ .70 indicates that all observed variables consistently measure the underlying factor, suggesting reliable subscale scores (Hair *et al*., 2014).

Validity refers to the degree to which a set of items measures the construct it intends to measure, meaning that the items accurately measure the construct rather than directly or indirectly measuring some other construct (Kline, 2016). Construct validity of scale scores can be established in terms of convergent and discriminant validity. Convergent validity (CV) refers to the degree to which a set of indicators converge properly by sharing a high proportion of common variance to explain the underlying dimension of the construct (Hair *et al*., 2019; Kline, 2016). CV describes how multiple indicators accurately and appropriately represent the underlying unobserved construct by eliminating unreliable indicators. In short, the latent variable must be adequately explained by its corresponding observed variables (Gefen *et al*., 2000).

Convergent validity can be assessed by the Average Variance Extracted (AVE) index, which indicates how much of the variance of the indicator variables is explained by an underlying factor with the criteria of statistically significant ($p < .000$) and adequate (≥ .50 or ideally ≥ .70) standardized factor loadings (Cheung *et al*., 2023; Hair *et al*., 2014). AVE is the average of all squared regression weights for observed variables belonging to the respective latent variable and is computed by the sum of squared standardized factor loadings divided by the sum of squared standardized factor loadings plus the sum of error variances (Hair *et al*., 2014). Additionally, internal consistency of the factor can also be considered to ensure the CV of the scale, as it assesses the degree to which all indicators measuring the same construct are correlated.

Discriminant Validity (DV), on the other hand, refers to how the observed variables of a parent factor distinctly measure that particular factor without displaying high correlations with other factors. Each factor should be well explained by its respective set of indicator variables. For instance, items related to the SECRP dimension should not significantly share variance with items from other dimensions, such as SERA, SEPA, and SEPB. DV is typically assessed using the Fornell-Larcker Criterion (FL) and the Heterotrait-Monotrait (HTMT) ratio of correlations (Cheung *et al*., 2024; Henseler *et al*., 2015). According to the F-L Criterion, if the AVE exceeds the squared correlations of the inter-factor correlations (or if the square root of the AVE for each factor is greater than its correlation with any other factor), then DV is supported.

3.3.1. *Establishing the reliability of the TICMSE scale*

The reliability analysis results for the 25-item TICMSE scale indicate that Cronbach's alpha values were .905, .902, .913, and .855, while McDonald's omega values were .907, .905, .916, and .856 for the SECRP, SEPA, SERA, and SEPB dimensions, respectively.

Table 3. *Descriptive statistics, reliability, and validity measures for the dimensions of the TICMSE scale*

Domain	Item Numbers	N	M	SD	AVE	CR	Cronbach's α	McDonald's ω
SECRP	1, 2, 4, 5, 6, 7	06	4.11	.813	.62	.91	.905	.907
SEPA	1, 2, 3, 4, 5, 6, 7	07	3.70	.785	.56	.90	.902	.905
SERA	1, 2, 3, 5, 6, 7	06	3.98	.833	.63	.91	.913	.916
SEPB	1, 2, 3, 5, 6, 7	06	4.05	.573	.50	.86	.855	.856
Scale	--	25	3.95	.540	--	--	.920	.921

Note: SECRP: Self-efficacy for enforcing classroom rules and procedures; SEPA: Self-Efficacy to adopt Proactive Approaches; SERA: Self-Efficacy for adopting Reactive Approaches; SEPB: Self-Efficacy to improve Prosocial Behavior; M: Mean; SD: Standard Deviation; AVE: Average Variance Extracted; CR: Composite Reliability.

These values surpass the cutoff value of .80, indicating strong internal consistency of the dimensions. The CR results were .91,.90,.91, and .86 for SECRP, SEPA, SERA and SEPB, respectively. Both measures affirm the internal consistency of the scale.

3.3.2. *Establishing the validity of the TICMSE scale*

The results of the CV analysis presented in Table 3 indicate that the AVE values for the SECRP, SEPA, SERA, and SEPB factors were .62, .56, .63, and .50, respectively, all of which exceed the .50 threshold (Hair *et al*., 2014). This suggests that the items within each factor converged adequately. In the factor correlation matrix shown in Table 4, the square root of the AVE for each factor surpasses the inter-factor correlations. Since all factors meet this criterion, the CV is thus established.

Although the F-L criterion has been extensively utilized in prior research on DV, it has faced criticism in recent discussions and developments (Henseler *et al*., 2015). Therefore, the present study has opted for alternative methods of assessing DV, such as the HTMT ratio of correlations. This approach examines whether the latent variables in a measurement model have stronger correlations with their observed variables by comparing items from other latent variables, thereby comparing correlations between latent variables (heterotrait correlations) to correlations between observed variables (monotrait correlations). Henseler *et al*. (2015) recommend a critical value for the HTMT of ≤ .85. Since all values displayed in Table 5 are below .85, indicating no issues in the discriminative power of the constructs, the presence of DV is confirmed.

Table 4. *Inter-Factor correlations with square roots of AVE for each factor.*

Factors	SECRP	SEPA	SERA	SEPB
SECRP	*0.787*			
SEPA	0.386	*0.748*		
SERA	0.383	0.492	*0.793*	
SEPB	0.385	0.422	0.387	*0.707*

Note: The diagonal contains the square root of AVE and off-diagonal cells contain the inter-factor correlations. SECRP: Self-efficacy for enforcing classroom rules and procedures; SEPA: Self-Efficacy to adopt Proactive Approaches; SERA: Self-Efficacy for adopting Reactive Approaches; SEPB: Self-Efficacy to improve Prosocial Behavior.

Table 5. *Correlation matrix showing the HTMT ratio of correlations for the factors.*

	SECRP	SERA	SEPA	SEPB
SECRP	---			
SERA	.410	---		
SEPA	.422	.342	---	
SEPB	.373	.338	.378	---

Note: Off-diagonal cells contain the HTMT ratio of correlations between different factors. SECRP: Self-efficacy for enforcing classroom rules and procedures; SEPA: Self-Efficacy to adopt Proactive Approaches; SERA: Self-Efficacy to adopt Reactive Approaches; SEPB: Self-Efficacy to improve Prosocial Behavior.

3.4. Descriptive Statistics for the TICMSE Scale Scores

Descriptive statistics for the TICMSE scale scores (Table 3) showed that the level of self-efficacy in inclusive CM of preservice teachers was above average for the overall ($\bar{x}$ = 3.95, SD =.540) as well as in the subscales, indicating that teachers were more likely to believe in their abilities to implement CM practices in their inclusive classrooms. Additionally, the SECRP subscale had the highest mean scores ($\bar{x}$ = 4.11, SD = .813), while the SEPA had the lowest ($\bar{x}$ = 3.70, SD = .785), which indicates that those who believe in their abilities to enforce rules, regulations, and procedures to discipline their students are less likely to adopt proactive approaches related to CM, consistent with the study of Doyle (1986), where he asserted that when teachers take on the role of disciplinarians, they are more likely to enforce reactive actions such as reprimands, punishments, etc., rather than prevent problematic behaviors. Also, the reduced standard deviation for the SEPB subscale scores (SD =.573) indicates that participants had chosen values closer to the mid-range of the response options, indicating less confidence in selecting extreme values, which implies that they were less likely to adopt practices to improve students' prosocial behavior compared to other strategies in their classrooms.

4. DISCUSSION and CONCLUSION

Despite the growing recognition of the significant contribution that ICMSE plays in promoting equitable learning opportunities for learners with SEN, the analysis of existing instruments has revealed a noticeable gap in the availability of standalone instruments specifically designed to assess this construct. Moreover, the available measures are deemed to estimate the unidimensional nature of the construct by encompassing a narrower range of CM efforts rather than incorporating different facets of the CMSE construct. Above all, the existing self-efficacy measures have various problems with conceptualizations and psychometric properties (O'Neill & Stephenson, 2011; Tschannen-Moran & Woolfolk Hoy, 2001). To address this gap, therefore, the present study aimed to investigate the dimensionality of the author-developed TICMSE scale and to assess its psychometric properties using EFA and CFA to systematically capture and quantify the beliefs of preservice teachers about their abilities to implement CM practices in inclusive settings within the Sri Lankan context.

Two main data reduction/extraction techniques, PCA and common FA (EFA), are widely used as multivariate statistical approaches in theory and for the development of psychometrically sound instruments, as well as for the refinement, evaluation, and validation of existing instruments in various fields, including Education, Psychology, and Management (Briggs & Cheek, 1986). Although these two techniques differ in purpose and computation, both serve the purpose of reducing or summarizing a large number of observed variables into relevant components (as in PCA) or factors (as in EFA) (Samuels, 2017). Although PCA is referred to and used as a method in data reduction, the principal components extracted are not ideally latent factors and do not equate with common factors (Fabrigar *et al*., 1999; Osborne, 2014). EFA allows for a more nuanced exploration of the underlying structure of the observed data. Therefore, this study employs EFA to examine the TICMSE scale's underlying structure and ensure that all items in the scale are valid and reliable indicators of the construct being researched.

The analysis of the TICMSE scale began with 40 items, with 10 items representing each factor. After removing four items during pretesting and six during content validation due to redundancy, the remaining 30 items were subjected to EFA. PAF with the DO rotation method yielded four separate but correlated factors, explaining 52.7% of the variance in the data. Most items were loaded onto their respective factors, the scale was further reduced to 27 items by discarding three items (SEPB4, SEPB8, and SERA4) that exhibited weak, as well as cross-loadings. The four factors, consisting of 27 items in total, accounted for 56.4% of the variance in respondents' scores. These factors were labelled as SECRP, SERA, SEPA, and SEPB, aligning well with the theoretical dimensions of the construct of ICMSE. Due to the complexity

of the construct under study and the substantial number of items in each factor, the decision was made to retain these four factors, despite their variance being slightly lower, but close to 60%, which is theoretically meaningful, interpretable, and satisfactory in the field of social sciences (Hair *et al*., 2014).

The four-factor structure of the scale extracted from EFA was statistically confirmed by CFA, with the number of items reduced from 27 to 25. The CFA revealed significantly weighted factor loadings, accounting for more than 50% of the total variance by the extracted factors, except for items SECRP7, SEPB3, and SEPB5, which showed standardized factor loadings ≤ .70. Although these values were slightly below the cutoff, they were very close it, and as they did not affect the overall model fit in CFA therefore, the decision was made to retain these items in the final scale for interpretation and further use. The high-weighted standardized lambda further confirmed the construct validity of the dimensions of the construct.

Regarding reliability measures, the literature recommends McDonald's ω over Cronbach's α for estimating internal consistency due to several limitations associated with α. First, Cronbach's α assumes normality of the data distribution and is sensitive to high skewness. Second, it may yield biased estimates in the presence of multidimensionality and correlated measurement errors. Third, α is affected by heterogeneous item correlations, particularly when factor loadings vary across items (Cheung *et al*., 2024). Given these concerns, McDonald's ω was used to interpret the internal consistency of the scale in this study (Malkewitz *et al*., 2023).

Reliability analysis using McDonald's omega estimates showed that all 25 items in the final scale had internal consistency, indicating that all items function in the same direction in measuring the underlying latent construct of ICMSE. Furthermore, all dimensions of the scale except for the SEPB subscale showed good internal consistency with values > .90. Because the SEPB subscale had two items (SEPB3 and SEPB5) with variance less than but closer to 50% (with low SD = .540), it had a low coefficient α and ω compared to other dimensions. Again, the same reason can be attributed to the other validity and reliability estimates, including AVE and CR, where the values were lower than those of other subscales.

The literature reviewed reveals a progressive evolution in the measurement of CMSE, transitioning from reliance on a single-item measure to the development of multi-item scales (Tschannen-Moran & Woolfolk Hoy, 2001). This shift is evident in instruments, such as the revised Dutch Teacher Self-Efficacy Scale (DTSES) by Meijer and Foster (1988), the Classroom and School Context teacher efficacy scale by Friedman and Kass (2002), and the Culturally Responsive Teaching Self-Efficacy (CRTSE) scale by Siwatu (2007). Notably, many items included in these scales, particularly in CRTSE, capture a broader conceptualization of CMSE, encompassing both reactive and proactive strategies, as well as efforts to actively engage students in learning and socialization, rather than focusing solely on disciplinary actions. This broader framing aligns with the conceptualization of CM adopted in the present study. However, it is worth noting that items specifically addressing classroom rules and routines were not prominent in those scales.

Additionally, the literature indicates that many published multidimensional SE scales include a distinct subscale for CMSE. Examples include the Dutch version of the Teacher Interpersonal Self-Efficacy (DTISE) scale, developed by Brouwers and Tomic (2001), the TSES by Tschannen-Moran and Woolfolk Hoy (2001), and the Norwegian Teacher Self-Efficacy Scale (NTSES) by Skaalvik and Skaalvik (2007). The DTISE scale features a 14-item subscale on Perceived Self-Efficacy in Classroom Management, which was adapted from Emmer and Hickman's earlier work (Brouwers & Tomic, 2001).

Prior to influential work of Tschannen-Moran and Woolfolk Hoy (2001), most SE studies relied on the adapted version of the 36-item Teacher Efficacy in Classroom Management and Discipline (TECMD) scale developed by Emmer and Hickman (Tschannen-Moran & Woolfolk Hoy, 2001). This was one of the first instruments to introduce a specific factor on CM and

discipline self-efficacy, comprising 14 items. Subsequently, the 24-item TSES by Tschannen-Moran and Woolfolk Hoy included a 9-item Efficacy for CM subscale. Exploratory factor analysis using PAF with varimax rotation yielded a three-factor structure, with loadings ranging from .50 to .78, accounting for 51.72% of the total variance, and a Cronbach's alpha of .90 for the CM subscale. Similarly, the NTSES developed by Skaalvik and Skaalvik (2007) consists of 24 items across six factors, including a four- item discipline subscale. This dimension focuses on teachers' ability to enforce classroom rules and maintain order. Factor analysis using the maximum likelihood method and varimax rotation produced factor loadings between .67 and .81, with a Cronbach's alpha of .90 for the discipline dimension.

On the other hand, a few subsequent instruments, such as the 14-item Behavior Management Self-Efficacy scale (BMSES) developed by Main and Hammond (2008), adapted from Baker (2005), which itself adapted from Brouwers and Tomic (1999), the 4-item Teacher-perceived Self-Efficacy in Classroom Management scale by Betroret (2009), and the 14-item Classroom Management Self Efficacy Instrument (CMSEI) by Slater and Main (2020) solely measure the CMSE construct. The unidimensional scale by Betroret, extracted by PCA with varimax rotation, yielded a high total variance of 72.57% explained by the factor, with factor loadings > .50.

Two strands of argument can be made to explain the inconsistencies between the factor structure of the BMSES and the TICMSE scale used in the present study. First, the BMSES specifically targets teachers' efficacy in managing students' behavior in the classroom, which represents only one component of the broader CM construct (Sciuchetti & Yssel, 2019). Second, although the BMSES encompasses a range of actions related to behavior management, such as implementing classroom rules, routines, and procedures, engaging students in learning, and maintaining order and control, similar to the domains addressed in the TIMCSE scale (excluding the socialization of students), its structure was reported as unidimensional. However, the original study provided limited evidence to substantiate this unidimensionality or to establish broader psychometric properties beyond internal consistency, which was reported with a Cronbach's alpha of .881. In constrst, the psychometric properties of the BMSES were evaluated through Rasch analysis by Slater and Main (2020) for the CMSEI, which comprised the same items as the BMSES originally developed by Main and Hammond (2008).

The following instruments, identified in the reviewed literature, are deemed appropriate for comparison with the TICMSE scale as they have demonstrated the multidimensionality of the construct of CMSE. The 22-item, two-factor, Efficacy for Classroom Management scale by Putman (2013). These two factors extracted using the PAF and DO rotation were labelled, general management efficacy (n = 15) and social comparisons for management (n = 7), with a Cronbach's alpha value of .94 for the total scale. The scale had a convergent validity of .61 (p < .01), established with the Efficacy for Classroom Management subscale of the TSES by Tschannen-Moran and Woolfolk Hoy (2001).

Overall, the analysis of the reviewed scales indicated that while CMSE is an important construct to investigate, many inconsistencies were observed, particularly in conceptualizing the construct. In several instruments, it is noted that the items did not capture all the nuances of effective CM. Some aspects of CM, such as establishing routines, preventing student misbehavior, and creating a positive learning environment by involving students in learning tasks, may not be fully captured (O'Neill & Stephenson, 2011). Moreover, many items included in the scales were not precisely aligned with Bandura's (1977) self-efficacy theory, as they seemed to assess teachers' general abilities, for example, "I can effectively manage classroom/my students", rather than their abilities to perform a specific task- domain CM (Tschannen-Moran *et al*., 1998). Additionally, several items with poor factor loadings (< .50) were included in the final scales, which were deemed irrelevant as they did not contribute significantly to the variations in the latent factors. Also, rigorous factor analysis techniques

were missing in establishing the factor structure and psychometric properties of many of the scales. Therefore, the TICMSE scale can significantly contribute to addressing these issues.

The vast majority of studies on TSE have highlighted the different dimensions of this construct, including CMSE, however, only a handful of studies have focused on various facets of the CMSE construct. For Lazarides *et al.* (2020), CM is an umbrella term that not only encompasses teachers' reactions to students' misbehavior but rather includes a range of efforts teachers make to establish an environment conducive to learning and the social-emotional development of students. In line with this, by capturing the four distinct dimensions of CMSE, this study ascertains the multidimensionality of the author-developed TICMSE scale. The identified dimensions, SERA, SEPA, SEPB and SECRP, considered mutually exclusive, have been shown to have excellent psychometric properties to help assess preservice teachers' beliefs regarding their abilities to establish and maintain a supportive inclusive learning environment for students.

4.1. Limitations and Further Research

As there is no strict rule in selecting the rotational techniques between orthogonal and oblique, which differ in the aspect that factors are to be correlated or not, this study utilized DO as the oblique rotation to allow factors to correlate (Hair *et al*., 2010). In the already published scales assessing the construct of ICMSE using subscales, the orthogonal technique is appropriate to confirm that the factors are uncorrelated for subsequent use in other multivariate approaches; however, in the present study, the use of DO is justifiable as the multidimensional TICMSE scale is an isolated measure where the factors should at least be correlated with each other. Nevertheless, the obliquely rotated factors have limitations in applying the findings to the different samples (Hair *et al*., 2010). Also, as validity and reliability are the properties of the scale scores, not the scale per se, the psychometric properties of the TICMSE scale should be established for its every application. The study was limited by gender bias due to the non-representation of male participants. Therefore, it is recommended that the study be extended by including male participants to assess whether there are any gender differences in terms of their ICMSE. Although this study, through rigorous analysis, established the psychometric properties of the original Tamil version of the ICMSE scale, the English version must be adapted to ensure whether there are linguistic and contextual differences that may influence how participants interpret the items in English so that to make the English version accessible for non-Tamil speaking contexts. Additionally, further study is needed to investigate the possible effects of teacher demographics, such as gender, nationality, and subject specializations, on the construct of ICMSE.

Declaration of Conflicting Interests and Ethics

The author declares no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the author(s). **Ethics Committee Number**: Faculty of Graduate Studies, University of Colombo, 16th December 2021, Ref No: FGS/ERC/2021/040.

Orcid

Yogaranee Sakthivel https://orcid.org/0009-0000-7803-3206

REFERENCES

Aas, H.K., Uthus, M., & Løhre, A. (2024). Inclusive education for students with challenging behaviour: Development of teachers' beliefs and ideas for adaptations through lesson study. *European Journal of Special Needs Education, 39*(1), 64-78. https://doi.org/10.1080/08856257.2023.2191107

Ainscow, M., & Miles, S. (2009). Developing inclusive education systems: How can we move policies forward? In C. Giné, J.M. Suriá, & M. Belda (Eds.), Educación inclusiva: De la

exclusión a la plena participación de todo el alumnado [Inclusive education: From exclusion to full participation of all students] (pp. 13–23). Universitat de València.

Alter, P., & Haydon, T. (2017). Characteristics of effective classroom rules: A review of the literature. *Teacher Education and Special Education, 40*(2), 114-127. https://doi.org/10.1177/0888406417700962

Alter, P., Walker, J., & Landers, E. (2013). Teachers' perceptions of students' challenging behavior and the impact of teacher demographics. *Education and Treatment of Children, 36*(4), 51–70. https://doi.org/10.1353/etc.2013.0040

Armor, D., Conry-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, A., Pauly, E., & Zellman, G. (1976). *Analysis of the school preferred reading program in selected Los Angeles minority schools* (Report No. R-2007-LAUSD). RAND Corporation. https://www.rand.org/pubs/reports/R2007.html

Ashton, P.T., Webb, R., & Doda, N. (1983). *A study of teachers' sense of efficacy* (Report No. ED231833). U.S. Department of Education, National Institute of Education, Educational Resources Information Center.

Ayre, C., & Scally, A.J. (2014). Critical values for Lawshe's content validity ratio. *Measurement and Evaluation in Counseling and Development, 47*(1), 79-86. https://doi.org/10.1177/0748175613513808

Baker, P.H. (2005). Managing student behaviour: How ready are teachers to meet the challenge? *American Secondary Education, 33*(3), 51–64.

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

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Bandura, A. (n.d.). *Teacher self-efficacy scale*. https://bit.ly/3HaBWIU

Bentler, P.M., & Bonnet, D.C. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606. https://doi.org/10.1037/0033-2909.88.3.588

Bentler, P.M., & Chou, C.P. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*(1), 78–117. https://doi.org/10.1177/0049124187016001004

Betoret, F.D. (2009). Self-efficacy, school resources, job stressors and burnout among Spanish primary and secondary school teachers: A structural equation approach. *Educational Psychology, 29*, 45–68. https://doi.org/10.1080/01443410802459234

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

Briggs, N.E., & MacCallum, R.C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research, 38*(1), 25–56. https://doi.org/10.1207/S15327906MBR3801_2

Briggs, S.R., & Cheek, J.M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality, 54*(1), 106-148. https://doi.org/10.1111/j.1467-6494.1986.tb00391.x

Brophy, J. (1982). *Classroom organization and management*. The Institute for Research on Teaching, Michigan State University.

Brophy, J.E. (2006). History of research in classroom management. In C.M. Evertson & C.S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 17–43). Lawrence Erlbaum Associates, Inc.

Brouwers, A., & Tomic, W. (1999). *The factorial validity of the teacher interpersonal self-efficacy scale* (No. SP038950) [ERIC Document Reproduction Service No. ED437365]. Netherlands Open University. https://eric.ed.gov/?id=ED437365

Brouwers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education, 16*(2), 239–253. https://doi.org/10.1016/S0742-051X(99)00057-8

Brouwers, A., & Tomic, W. (2001). The factorial validity of scores on the teacher interpersonal self-efficacy scale. *Educational and Psychological Measurement, 61*(3), 433–445. https://doi.org/10.1177/00131640121971356

Brown, T.A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.

Byrne, B.M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Routledge, Taylor & Francis Group.

Cheung, G.W., Cooper-Thomas, H.D., Lau, R.S., & Wang, L.C. (2024). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. *Asia Pacific Journal of Management, 41*(3), 745-783. https://doi.org/10.1007/s10490-023-09871-y

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). Routledge.

Costello, A.B., & Osborne, J.W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation, 10*(7), 1–9. https://doi.org/10.7275/jyj1-4868

Curran, P.J., West, S.G., & Finch, J.F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*(1), 16–29. https://doi.org/10.1037/1082-989X.1.1.16

De Silva, P.V., Hewage, C.G., & Fonseka, P. (2015). Prevalence of burnout and its correlates among female primary school teachers in the Southern Province of Sri Lanka. *European Journal of Preventive Medicine, 3*(2-1), 9-14. https://doi.org/10.11648/j.ejpm.s.2015030201.13

DeVellis, R.F. (2017). *Scale development: Theory and application* (4th ed.). Sage Publications.

Dicke, T., Parker, P.D., Marsh, H.W., Kunter, M., Schmeck, A., & Leutner, D. (2014). Self-efficacy in classroom management, classroom disturbances, and emotional exhaustion: A moderated mediation analysis of teacher candidates. *Journal of Educational Psychology*. Advance online publication. https://doi.org/10.1037/a0035504

Doyle, W. (1986). Classroom organization and management. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 392–431). Macmillan.

Eisenman, G., Edwards, S., & Cushman, C.A. (2015). Bringing reality to classroom management in teacher education. *Professional Educator, 39*(1). https://eric.ed.gov/?id=EJ1062280

Emmer, E., Evertson, C., & Anderson, L. (1980). Effective classroom management at the beginning of the school year. *The Elementary School Journal, 80*(5), 219-231. https://doi.org/10.1086/461192

Evertson, C., & Poole, I. (2008). Proactive classroom management. In T. Good (Ed.), *21st century education: A reference handbook* (pp. I-131–I-140). SAGE Publications. https://doi.org/10.4135/9781412964012.n14

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). SAGE Publications.

Friedman, I.A., & Kass, E. (2002). Teacher self-efficacy: A classroom-organization conceptualization. *Teaching and Teacher Education, 18*(6), 675-686. https://doi.org/10.1016/S0742-051X(02)00027-6

Garrote, A., Felder, F., Krähenmann, H., Schnepel, S., Sermier Dessemontet, R., & Moser Opitz, E. (2020). Social acceptance in inclusive classrooms: The role of teacher attitudes

toward inclusion and classroom management. *Frontiers in Education, 5*, Article 582873. https://doi.org/10.3389/feduc.2020.582873

Gefen, D., Straub, D., & Boudreau, M. (2000). Structural equation modeling and regression: Guidelines for research practice. *Communications of the Association for Information Systems, 4*, 1–78. https://doi.org/10.17705/1CAIS.00407

Goss, P., Sonnemann, J., & Griffiths, K. (2017). *Engaging students: Creating classrooms that improve learning*. Grattan Institute. https://bit.ly/3EXn2VX

Guadagnoli, E., & Velicer, W.F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin, 103*(2), 265–275. https://doi.org/10.1037/0033-2909.103.2.265

Güner-Yıldız, N., Köse, H., & Akın, E. (2022). Classroom management: Significance in inclusive education, current problems and proposed solutions. *Ankara University Faculty of Educational Sciences Journal of Special Education, 23*(3), 699-720. https://doi.org/10.21565/ozelegitimdergisi.856085

Guo, Y., Connor, C.M., Yang, Y., Roehrig, A.D., & Morrison, F.J. (2012). The effects of teacher qualification, teacher self-efficacy, and classroom practices on fifth graders' literacy outcomes. *The Elementary School Journal, 113*(1), 3–24. https://doi.org/10.1086/665816

Guskey, T.R. (1986, April 16–20). *Context variables that affect measures of teacher efficacy* [Paper presentation]. Annual Meeting of the American Educational Research Association, San Francisco, CA, United States. https://eric.ed.gov/?id=ED269398

Guskey, T.R., & Passaro, P.D. (1994). Teacher efficacy: A study of construct dimensions. *American Educational Research Journal, 31*(3), 627-643. https://doi.org/10.3102/00028312031003627

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2010). *Multivariate data analysis* (7th ed.). Pearson Education Limited.

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2014). *Multivariate data analysis* (7th ed., Pearson New International Edition). Pearson Education Limited.

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

Hair, J.F., Jr., Hult, G.T.M., Ringle, C.M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)* (2nd ed.). SAGE Publications.

Henseler, J., Ringle, C.M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*, 115–135. https://doi.org/10.1007/s11747-014-0403-8

Hooper, D., Coughlan, J., & Mullen, M.R. (2008). Structural equation modelling: Guidelines for determining model fit. *Journal of Business Research Methods, 6*, 53–60.

Hoy, W.K., & Woolfolk, A.E. (1993). Teachers' sense of efficacy and the organizational health of schools. *The Elementary School Journal, 93*(4), 355–372. https://doi.org/10.1086/461729

Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Kaiser, H.F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31–36. https://doi.org/10.1007/BF02291575

Kaplan, A., Gheen, M., & Midgley, C. (2002). Classroom goal structure and student disruptive behaviour. *British Journal of Educational Psychology, 72*(1), 191-211. https://doi.org/10.1348/000709902158847

Kline, R.B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research, 86*(3), 643–680. https://doi.org/10.3102/0034654315626799

Korpershoek, H., Mouw, J., & de Boer, H. (2022). International research on the foci and effectiveness of classroom management programs and strategies. In E.J. Sabornie & D.L. Espelage (Eds.), *Handbook of classroom management* (3rd ed., pp. 350–372). Routledge. https://doi.org/10.4324/9781003275312-22

Kutner, M.H., Nachtsheim, C.J., Neter, J., & Li, W. (2004). *Applied linear statistical models* (5th ed.). McGraw-Hill.

Lamm, K.W., Lamm, A.J., & Edgar, D. (2020). Scale development and validation: Methodology and recommendations. *Journal of International Agricultural and Extension Education, 27*(2), 24–35. https://doi.org/10.4148/2831-5960.1115

Landis, J.R., & Koch, G.G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics, 33*, 363–374. https://doi.org/10.2307/2529786

Lawshe, C.H. (1975). Content validity. *Personnel Psychology, 28*, 563-575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Lazarides, R., Watt, H.M.G., & Richardson, P.W. (2020). Teachers' classroom management self-efficacy, perceived classroom management, and teaching contexts from beginning until mid-career. *Learning and Instruction, 69*, 101346. https://doi.org/10.1016/j.learninstruc.2020.101346

MacCallum, R.C., Widaman, K.F., Preacher, K.J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research, 36*(4), 611–637. https://doi.org/10.1207/S15327906MBR3604_06

Main, S., & Hammond, L. (2008). Best practice or most practiced? Pre-service teachers' beliefs about effective behaviour management strategies and reported self-efficacy. *Australian Journal of Teacher Education, 33*(4). https://doi.org/10.14221/ajte.2008v33n4.3

Malkewitz, C.P., Schwall, P., Meesters, C., & Hardt, J. (2023). Estimating reliability: A comparison of Cronbach's α, McDonald's ω_t, and the greatest lower bound. *Social Sciences & Humanities Open, 7*, 100368. https://doi.org/10.1016/j.ssaho.2022.100368

Marzano, R.J., Marzano, J.S., & Pickering, D.J. (2003). *Classroom management that works: Research-based strategies for every teacher*. ASCD.

Matas-Terrón, A., Aranda, L., Franco-Caballero, P.D., & Mena-Rodríguez, E. (2024). Attitudes towards research methods in education: Development of the ATRMQ scale. *Education Sciences, 14*(374). https://doi.org/10.3390/educsci14040374

Meijer, C.J.W., & Foster, S.F. (1988). The effect of teacher self-efficacy on referral chance. *Journal of Special Education, 22*, 378–385. https://doi.org/10.1177/002246698802200309

Milner, H.R., & Tenore, F.B. (2010). Classroom management in diverse classrooms. *Urban Education, 45*(5), 560–603. https://doi.org/10.1177/0042085910377290

Oliver, R.M., & Reschly, D.J. (2007). Effective classroom management: Teacher preparation and professional development. In E.J. Sabornie & D.L. Espelage (Eds.), *Handbook of classroom management* (pp. 127–150). Lawrence Erlbaum Associates.

Oliver, R.M., Wehby, J.H., & Reschly, D.J. (2011). Teacher classroom management practices: Effects on disruptive or aggressive student behavior. *Campbell Systematic Reviews*, *4*. https://doi.org/10.4073/csr.2011.4

O'Neill, S.C. (2015). Preparing preservice teachers for inclusive classrooms: Does completing coursework on managing challenging behaviours increase their classroom management sense of efficacy? *Australasian Journal of Special Education*. Advance online publication. https://doi.org/10.1017/jse.2015.10

O'Neill, S.C., & Stephenson, J. (2011). The measurement of classroom management self-efficacy: A review of measurement instrument development and influences. *Educational Psychology, 31*(3), 261–299. https://doi.org/10.1080/01443410.2010.545344

Osborne, J.W. (2014). *Best practices in exploratory factor analysis*. CreateSpace Independent Publishing. ISBN-13: 978-1500594343, ISBN-10:1500594342.

Page, A., & Jones, M. (2018). Rethinking teacher education for classroom behaviour management: Investigation of an alternative model using an online professional experience in an Australian university. *Australian Journal of Teacher Education, 43*(11), 84–104. https://doi.org/10.14221/ajte.2018v43n11.5

Peters, J.H. (2012). Are they ready? Final year pre-service teachers' learning about managing student behaviour. *Australian Journal of Teacher Education, 37*(9), 18-42. https://doi.org/10.14221/ajte.2012v37n9.2

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

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

Putman, S.M. (2013). Measuring practicing and prospective elementary teachers' beliefs: Development and validation of the efficacy for classroom management scale. *Alberta Journal of Educational Research, 59*(3), 420-441. https://doi.org/10.55016/ojs/ajer.v59i3.55732

Rotter, J.B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied, 80*(1), 1-28. https://doi.org/10.1037/h0092976

Samuels, P. (2017). *Advice on exploratory factor analysis*. ResearchGate. https://bit.ly/3S9zWmK

Sciuchetti, M.B., & Yssel, N. (2019). The development of preservice teachers' self-efficacy for classroom and behavior management across multiple field experiences. *Australian Journal of Teacher Education, 44*(6), 1–21. https://doi.org/10.14221/ajte.2018v44n6.2

Shapiro, S.S., & Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3/4), 591–611. https://doi.org/10.2307/2333709

Sharma, U., Loreman, T., & Forlin, C. (2011). Developing the teacher efficacy for inclusive practices scale. *Teaching and Teacher Education, 27*(4), 734-743. https://doi.org/10.1037/t76434-000

Siwatu, K.O. (2007). Preservice teachers' culturally responsive teaching self-efficacy and outcome expectancy beliefs. *Teaching and Teacher Education, 23*, 1086-1101. https://doi.org/10.1016/j.tate.2006.07.011

Skaalvik, E.M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology, 99*, 611–625. https://doi.org/10.1037/0022-0663.99.3.611

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

Sokal, L., Woloshyn, D., & Funk-Unrau, S. (2013). How important is practicum to pre-service teacher development for inclusive teaching? Effects on efficacy in classroom management. *Alberta Journal of Educational Research, 59*(2), 285-298. https://doi.org/10.55016/ojs/ajer.v59i2.55680

Spector, P.E. (1992). *Summated scale construction: An introduction* (Vol. 82). Sage. https://doi.org/10.4135/9781412986038

Subban, P.K., & Round, P. (2015). Differentiated instruction at work: Reinforcing the art of classroom observation through the creation of a checklist for beginning and pre-service teachers. *Australian Journal of Teacher Education, 40*(5), 117-131. https://doi.org/10.14221/ajte.2015v40n5.7

Tabachnick, B.G., & Fidell, L.S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Taherdoost, H., Sahibuddin, S., & Jalaliyoon, N. (2014). Exploratory factor analysis: Concepts and theory. In J. Balicki (Ed.), *Advances in applied and pure mathematics* (pp. 375–382).

WSEAS. Mathematics and Computers in Science and Engineering Series. https://hal.archives-ouvertes.fr/hal-02557344

Tschannen-Moran, M., & Hoy, A.W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783–805. https://doi.org/10.1016/S0742-051X(01)00036-1

Tschannen-Moran, M., Hoy, A.W., & Hoy, W.K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*(2), 202-248. https://doi.org/10.3102/00346543068002202

Tucker, L.R., & MacCallum, R.C. (1997). *Exploratory factor analysis*. American Psychological Association.

UNESCO. (1994). *The world conference on special needs education: Access and quality. Final report*. Ministry of Education and Science, Madrid; UNESCO.

UNESCO. (2008). *Inclusive education: The way of the future*. International Conference on Education: Forty-eighth session, International Conference Centre, Geneva.

UNESCO. (2019). *Disability-inclusive education: Foundation level. Education learning and development module*. UNESCO.

Watkins, M.W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology, 44*(3), 219–246. https://doi.org/10.1177/0095798418771807

Webster-Stratton, C., Reinke, W.M., Herman, K.C., & Newcomer, L.L. (2011). The Incredible Years Teacher Classroom Management Training: The methods and principles that support fidelity of training delivery. *School Psychology Review, 40*(4), 509-529. https://doi.org/10.1080/02796015.2011.12087527

Wheaton, B., Muthen, B., Alwin, D.F., & Summers, G.F. (1977). Assessing reliability and stability in panel models. *Sociological Methodology, 8*(3), 84-136. https://doi.org/10.2307/270754

Wickramasinghe, E.P., Seneviratne, R.D.E.A., Gunawardana, N.S., Sumathipala, A., & Mallen, C.D. (2022). Prevalence of occupational stress among secondary school teachers in public schools in the district of Colombo, Sri Lanka. *Research in Health Science, 7*(3). https://doi.org/10.22158/rhs.v7n3p12

Wong, H.K., & Wong, R.T. (2009). *The first days of school: How to be an effective teacher*. Harry K. Wong Publications.

Woodcock, S., & Reupert, A. (2012). A cross-sectional study of student teachers' behaviour management strategies throughout their training years. *Australian Educational Researcher, 39*, 159–172. https://ro.uow.edu.au/edupapers/1081

Yogaranee, S. (2024). Perspectives of pre-service and in-service general education teachers of the Faculty of Education on inclusive education: A phenomenological study. *Research Journal of Educational Sciences, 12*(2), 1–14. https://www.isca.me