2020

A measure of classroom management: Validation of a pre-service teacher self-efficacy scale

Eileen Slater
Edith Cowan University

Susan Main
Edith Cowan University

10.1080/02607476.2020.1770579
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

Journal of Education for Teaching
International research and pedagogy

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cjet20

A measure of classroom management: validation of a pre-service teacher self-efficacy scale

Eileen V. Slater & Susan Main

JOURNAL OF EDUCATION FOR TEACHING
https://doi.org/10.1080/02607476.2020.1770579

A measure of classroom management: validation of a pre-service teacher self-efficacy scale

Eileen V. Slater and Susan Main

Edith Cowan University, School of Education, Mount Lawley, Western Australia

ABSTRACT

Classroom management skills are essential for effective teaching and consequently form an integral part of undergraduate teaching degrees. Self-efficacy in classroom management influences an individual's willingness to undertake specific actions and their perseverance in the face of difficulties in executing these actions. In order to track the progress of pre-service teachers' self-efficacy in classroom management, an easy to administer Classroom Management Self Efficacy Instrument (CMSEI) was developed and piloted with a third year cohort of pre-service teachers. This article reports on the psychometric properties of the CMSEI as determined through a Rasch analysis. The analysis supports the Classroom Management Self Efficacy Instrument (CMSEI) as an accurate and internally consistent, unidimensional scale for use with undergraduate pre-service teachers.

ARTICLE HISTORY
Received 19 September 2019
Accepted 14 May 2020

KEYWORDS
Classroom management; self-efficacy; pre-service teachers; Rasch

Introduction

Good classroom management is an integral part of effective teaching (Martella et al. 2012; Marzano, Marzano, and Pickering 2003; Postholm 2013). It has been reported that good classroom management positively affects student outcomes while poor classroom management results in loss of teaching and learning time and poor educational outcomes (Goss, Sonnemann, and Griffiths 2017; Jones and Jones 2012).

Effective classroom management includes elements of organisation, rule setting and enforcement, managing resources, gaining and maintaining student attention, monitoring task engagement and modelling and reinforcing appropriate social interactions (Marzano, Marzano, and Pickering 2003). Overall, the focus should be on preventive, rather than reactive, classroom management procedures (Lewis and Sugai 1999; Sailor et al. 2009) and definitions of effective classroom management emphasise the actions that the teacher undertakes to facilitate learning (Brophy 2006; Evertson and Weinstein 2006). Due to its importance, understanding and managing behaviour has a long history of research dating back to Thorndike (1919) and Skinner (1953). However, the 1970's saw the start of a number of large-scale, systematic studies of classroom management (Brophy and Evertson 1976; Kounin 1970) that continued through the 1980's and 1990's (Brophy 1996; Evertson and Weinstein 2011).

CONTACT Eileen V. Slater e.slater@ecu.edu.au

In the contemporary Australian context, classroom management techniques are based on a range of theoretical positions but are predominantly driven by humanistic philosophies such as those emanating from John Dewey (1899) and Rogers (1957). These can be seen in the psycho-educational models of classroom management that focus on teachers understanding inappropriate behaviour as a flawed attempt to deal with the demands of the learning environment. The role of the teacher is to help students appreciate the need for change and, therefore, make better behavioural choices. Other influences include an emphasis on prevention (Kounin 1970) and positive discipline (Rogers 2015). There is also a renewed acknowledgement that behaviourist approaches have something to contribute to understanding behaviour and supporting students to choose more appropriate behaviour. Teachers in this context will have been exposed to these approaches and should feel confident in applying them.

Classroom management self-efficacy

Self-efficacy is conceptualised as the individual's belief in their ability to undertake the actions necessary to successfully accomplish specific tasks in specific contexts (Bandura 1986). Understandings of teacher self-efficacy drawn from Tschannen-Moran, Hoy, and Hoy's (1998) model acknowledge the impact of Rotter's (1966) locus of control theory and Bandura's (1997) sources of influence for the construction of self-efficacy. Combined, these theories identify the significance of factors such as the individual's perceptions of the control they have over outcomes, opportunities to observe similar others successfully performing tasks, encouragement from others that they possess what is necessary to achieve success, experiences of success, and the inferences drawn from their physical and emotional states.

Bandura (1997) asserted that, when teachers have a strong sense of self-efficacy, they are more likely to perceive they have influence over student outcomes. However, the predictive value of self-efficacy has been questioned by some researchers who suggest that the relationship between self-efficacy and outcomes is dependent on individual factors including personality traits and the difficulty of the tasks being undertaking (Judge et al. 2007). The influence of the specific self-efficacy scale used and the amount of teaching experience of respondents have also been identified as factors in the relationship between self-efficacy and student outcomes (Kim and Seo 2018). Other concerns about the validity of using self-efficacy as an indicator of performance relate to the Dunning-Kruger effect. That is, individuals may have a strong sense of self-efficacy because they are not aware of what they do not know (Kruger and Dunning 1999).

Despite these concerns, numerous studies have provided support for an association between high self-efficacy and student achievement (Kim and Seo 2018; Talsma et al. 2018; Tschannen-Moran and Woolfolk-Hoy 2001). Self-efficacy has also been shown to be an imperative part of teacher general self-efficacy and workforce readiness (Dicke et al. 2014; Flower, McKenna, and Haring 2017; Korpershoek et al. 2016). In this study, the researchers were interested in self-efficacy in classroom management, a specific subset of self-efficacy that is defined as 'teachers' beliefs in their capabilities to organize and execute the courses of action required to maintain classroom order' (Brouwers and Tomic 2000, 242).

When teachers perceive themselves as capable of managing the classroom environment they are less stressed, less likely to experience 'burnout' (Aloe, Amo, and Shanahan 2014) and, consequently, more likely to remain in the teaching profession (Woodcock and Reupert 2012). The importance of self-efficacy in pre-service teacher education has also been noted and poor self-efficacy for classroom management has been attributed to the high level of attrition in early career teachers (Simonsen et al. 2014). Researchers have suggested that graduate teachers often feel unprepared when entering the teaching environment and hold concerns about their ability to manage the class (Page and Jones 2018; Subban and Round 2015). A recent study of 1,227 German pre-service teachers also found that low self-efficacy was a predictor of emotional exhaustion where classroom behaviour was challenging (Dicke et al. 2014) and emotional exhaustion has been linked to motivation to leave teaching (Skaalvik and Skaalvik 2017).

A study by Baker (2005), from which the instrument analysed in this research was developed, examined teacher beliefs about their self-efficacy in relation to general classroom management skill as well as readiness to implement specific behaviour management techniques. The survey was distributed to 885 primary and high school teachers in one region of the United States of America with 345 responses. The findings indicated that teachers with higher self-efficacy also reported greater readiness to manage challenging students.

Reflecting the importance of classroom management, classroom management self-efficacy (CMSE) items have gradually become more prevalent as sub-scales in general teaching SE scales (O'Neill and Stephenson 2011) and single scales have been developed targeting CMSE in isolation. In 2008, Main and Hammond developed a scale to measure Australian pre-service teachers' self-efficacy in classroom management. This scale was based on the *Teacher Readiness Scale for Managing Challenging Classroom Behaviours* (Baker 2005), which had drawn some of its items from the *Teacher Interpersonal Self-Efficacy Scale* (Brouwers and Tomic 2001). This new scale was piloted with a third year cohort of pre-service teachers to assess its reliability as a scale to measure the development of pre-service teacher self-efficacy during the undergraduate teacher education course, with the intent to inform course design. This work was initially reported by Main and Hammond (2008) as a pre-post design study, where Cronbach's coefficient alpha for the scale was 0.881.

Subsequently, in their review of CMSE scales, O'Neill and Stephenson (2011) identified that a key issue with the scales they reviewed was the lack of consideration on how to increase the reliability coefficients and called for publication of 'item influences, reliability and validity testing conducted, item FL [factor loading] scores, and sample scores for individual items as well as subscale scores would enhance the utility and confidence in SE research findings' (O'Neill and Stephenson 2011, 295). In addition to O'Neill and Stephenson (2011) conclusions that a more rigorous approach should be made to the assessment and reporting of validity and reliability evidence in relation to SE scales, Berg and Smith (2016) urged for further research on self-efficacy in the Asia-Pacific region, including Australia.

The authors in the Main and Hammond (2008) study received numerous requests from independent colleagues in Australia and America in relation to the specific psychometric properties of the scale they utilised, beyond the reported coefficient alpha. As a result of the combined evidence for a continued need for quality CMSE scales with sound

psychometrics, particularly in the Australian context, a Rasch analysis of Main and Hammond's scale was conducted at the request of the first author of that study.

Method

Instrument

O'Neill and Stephenson (2011) challenged researchers to design CMSE scales and subscales that more closely align with Bandura's (1997) construction of self-efficacy and include items about classroom management that challenge teachers (items that are more difficult to endorse). Self-efficacy is defined by Bandura as 'a judgment of capability to execute given types of performances' (2006 p. 309). He further elaborates that efficacy scales should be about what the respondent can do, not what they intend to do, and avoid including items that confuse self-efficacy with the constructs of self-esteem (what is my worth), locus of control (are outcomes within or outside my control), and outcome expectancies (what is likely to flow-on from a particular performance).

In the development of the Classroom Management Self Efficacy Instrument (CMSEI) analysed in this research, existing classroom management scales were reviewed in light of Bandura's recommendations. The *Teacher Readiness Scale for Managing Challenging Classroom Behaviours* (Baker 2005) was selected for further consideration as it contained items which were statements of ability 'I can' or 'I am able to', reflecting efficacy as a judgement of performance. In addition, good face and construct validity were reported and items in Baker's (2005) scale were reflective of current classroom management theory taught in pre-service teacher courses in Australia (Kim and Seo 2018). Baker's (2005) instrument was modified to include, theoretically, the least number of items that would still encompass the breadth of the construct. Items were removed that did not reflect the emphasis placed on humanistic and ecological approaches to classroom management that are priorities in the Australian context and are, therefore, reflected in classroom management units of study in pre-service teaching degrees. For example, items relating to conducting a functional behaviour assessment (FBA) were removed from the questionnaire as pre-service teachers generally do not learn how to conduct a FBA or have the opportunity to observe this process in their practicum experiences. Other items that were related to specific approaches not covered in general pre-service teacher education, such as life-space intervention, therapeutic holding and role play were also removed. In addition, as the respondents for this survey were pre-service teachers, items with specific relevance to in-service teachers were also removed.

In establishing face validity, feedback was sought from five professionals in the field of classroom management, three who were currently teaching classroom management in undergraduate and post graduate education courses at tertiary institutions and two who were qualified special educators with over 15 years teaching in schools. The scale was then piloted with five pre-service teachers from the third year cohort, with follow-up discussion, who reported no issues or confusion in interpreting the items on the questionnaire for the proposed scale. Figure 1 shows the final items piloted with third year pre-service teachers as the CMSEI, using a four point Likert scale ranging from 'strongly disagree' to 'strongly agree'.

1. I am able to use a variety of behaviour management models and techniques.
2. If a student disrupts the lesson, I am able to redirect him/her quickly.
3. I can communicate to students that I am serious about getting appropriate behaviour.
4. There are very few students that I cannot handle.
5. I can manage a class very well.
6. I can keep defiant students involved in my lessons.
7. I am able to make my expectations clear to my students.
8. I can keep a few problem students from ruining an entire class.
9. If students stop working, I can put them back on track.
10. I know what rules are appropriate for my students.
11. I am able to use a variety of non-aversive techniques (e.g., voice modulation, facial expressions, planned ignoring, and proximity control).
12. I am able to implement a consistent classroom routine.
13. I am able to self-evaluate my own teaching and classroom management skills and use the results constructively
14. I am able to explain the rationale, program components, operation, and evaluation of the behavioural techniques I use.

Figure 1. Final Classroom Management Self Efficacy Instrument (CMSEI) items.

In addition to the classroom management self-efficacy scale items, the questionnaire collected demographic data including: age, gender, programme, past experience working with children, and parental status, which it was considered may impact upon participants' classroom management self-efficacy.

Sample

As part of a broader study by Main and Hammond (2008), 302 pre-service teachers in their third year of a four year undergraduate education degree were invited to complete the CMSEI to determine their self-efficacy in classroom management. A total of 123 (41%) of the pre-service teachers responded to the pre-intervention questionnaire and 69 (23%) responded to the post intervention questionnaire: 18 males and 51 females. For the purpose of scale validation, 69 matched pre-post cases were used, resulting in a stacked sample for analysis totalling 138.

The third year cohort was targeted in the piloting of the CMSEI because they hypothetically sit at a midpoint in their classroom management skills development; having completed a classroom management unit in the second year of the undergraduate teaching degree and their third year, four week, practicum. Therefore, the targeting of the instrument for third year students can give a good indication of whether the instrument may have floor or ceiling effects, which has implications for the potential of the instrument to be used with less or more experienced undergraduates to track the construct across time.

Once ethics approval was obtained from the universities Human Research Ethics Committee, the researcher spoke to the pre-service teachers whilst they were attending a lecture in their core inclusive education unit. She explained the purpose of the research and the research procedures. At the time of the research the researcher was a sessional tutor in the unit in which students were surveyed and was undertaking this research as part of her Masters study. The pre-service teachers were informed, both during the lecture and in the information letter, that their decision on whether to participate would not influence their outcomes in the unit. Pre-service teachers indicated their consent to being involved in the research by completing the questionnaire and returning it to the researcher.

Data analysis

Rasch

Drawing on the data from the previous research, the psychometric properties of the CMSEI were investigated using the Rasch model (Rasch, 1961). Rasch is based on the principal that interval level data can be derived when increases in the level of an attribute are related to the difficulty of the question and the ability of the person (Bond and Fox 2015). Rasch analysis places persons and items on the same scale, measured in logits. The use of the logits scale allows for simple visualisation of the difficulty of items and the location of persons in relation to the items, the equal interval scale opens up a wide range of statistical analysis techniques and it allows for comparisons within groups of persons that are independent of the items chosen, and vice versa (Andrich and Styles 2004). The major benefit in educational contexts is that when students' scores are summed and totals are used to make educational judgements, the scores represent a true difference in performance.

Rasch (1960/1980) models of Modern Test Theory have increasingly been used to determine the psychometric properties of instruments in the fields of psychology, education and health (Cano, Barrett, Zajiceck and Hobart 2011; Hagquist, Bruce and Gustavsson 2009; Tennant and Connaghan, 2007). In the field of education, the Rasch model has been used for some time to analyse the responses of students in international testing programmes such as the Programme for International Student Assessment (PISA) and national testing programmes such as the Australian National Assessment Programme – Literacy and Numeracy (NAPLAN).

The software RUMM2030 (Andrich, Sheridan, and Luo 2014) was used to analyse responses to the instrument. Data were entered for matched cases for both pre and post assessment using Time One (T1, pre-intervention, before the coursework began) and Time Two (T2, post-intervention, at the completion of the coursework) as person variables. Therefore, the analysis represented 138 responses, which could be separated into comparison groups to consider the validity and reliability of the CMSEI for pre and post use. A sample of this size is adequate to pilot this instrument, with item calibrations and person estimates expected to be stable in this context to ± 0.5 logits (95% confidence). The recommended minimum sample size is 50 (Linacre, n.d.).

Person/item alignment and reliability

Cronbach's coefficient alpha as well as the Rasch index of reliability, called the Person Separation Index (Andrich 1982), were examined as measures of internal consistency and reliability respectively.

The Rasch analysis provides an estimate of the pre-service teachers 'difficulty to endorse' estimate for each item in the CMSEI scale, with positive logit values indicating items which were more difficult to endorse and persons with higher efficacy. Analysis of person-item histograms was used to consider the targeting of the items to the persons.

Data fit to the model

Rasch analysis identifies items that do not 'fit' with other items, that is, they don't measure the same construct. Fit was assessed statistically through two fit statistics, the Fit residual and Chi Square fit statistics, as well as graphically through the Item Characteristic Curves

(ICC). If the mean (*M*) and standard deviation (*SD*) of the fit residual fit statistic is close to 0 and 1 respectively it indicates that the data fit the model. If an individual item fits the model, it will have Fit residual values between −2.5 and 2.5. The Chi Square fit statistic compares observed mean responses with what is expected according to the Rasch model, and indicates misfit when the divergence is statistically significant. To avoid a type one error, the Bonferroni adjustment (Bland and Altman 1995) was applied to determine significance levels in this study.

Violations of independence

A principal component analysis (PCA) of residuals was made, where the absence of any meaningful pattern supports the assumption of unidimensionality. RUMM2030 (Andrich, Sheridan, and Luo 2014) also provides an estimate of the error adjusted correlation between the underlying traits measured by different groupings of items, reported as an A value. If a scale is unidimensional, a comparison of subtests should yield a similar PSI and a high A value.

Differential item functioning (DIF)

A differential item functioning (DIF) analysis identifies items that function differently for subgroups of the population. DIF was assessed graphically through an inspection of the Item Characteristic Curve (ICC) for each item and confirmed statistically through an ANOVA of the residuals.

Response category functioning

The Likert scale was assessed for a minimum number of responses per category of 10 (Linacre 1999), that the category thresholds progress monotonically (category thresholds represent the point at which the difficulty of endorsement of two adjacent categories has 50:50 probability), that the category probability graphs indicate monotonic progression of categories, and in the case of a four point scale, the distance between category thresholds is at least 1.4 logits and no more than 5 logits. The acceptable distributions for response categories are uniform, normal, bi-modal or slightly skewed (Bond and Fox 2015).

Results

Item/person alignment

Figure 2 shows histograms, on the same scale, of the Rasch person estimates (top histogram) and item difficulty estimates (bottom histogram) for the stacked sample (N = 138). The mean of the person estimates was 1.59 relative to the mean of items which is constrained to be 0, supporting the hypothesis that many participants found the items easy to endorse. The person separation index (PSI) was 0.89 indicating an excellent ability of these tests to detect misfit and good reliability, and coefficient alpha was 0.90, suggesting good internal consistency.

There is an absence of items to separate the persons above 1.6 logits, suggesting further investigation of the targeting of the instrument is warranted. Figure 3 shows the category threshold distribution which indicates a better targeting of the instrument to the persons with thresholds covering the full range of person abilities.

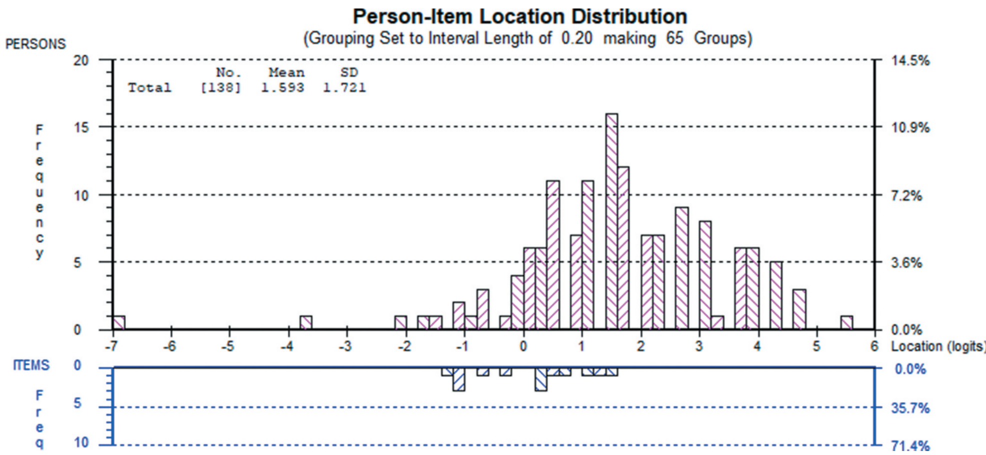

Figure 2. Person-item location distribution.

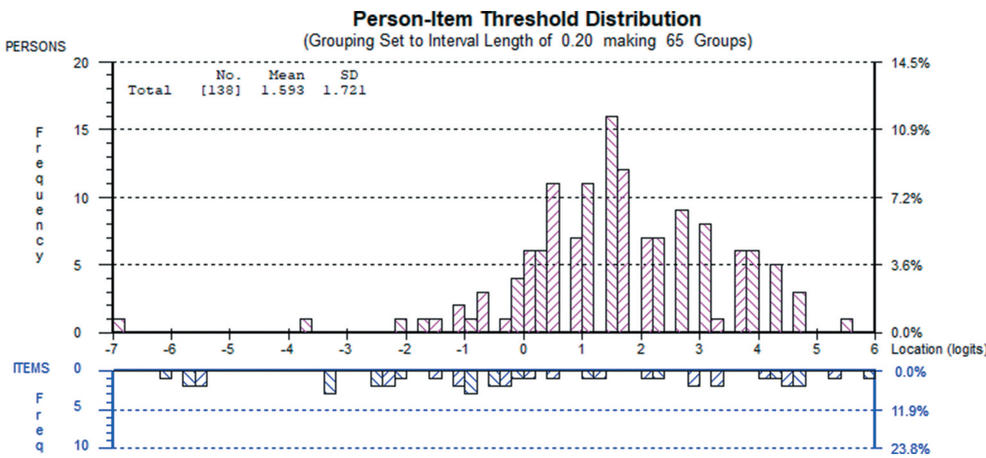

Figure 3. Person-item threshold distribution.

Data fit to the model

Table 1 shows the summary fit statistics for the pre, post and stacked sample, for the 14 item CMSEI, providing evidence of fit to a unidimensional model. In the stacked sample, the mean of the item fit residual fit statistic was 0.09 and the SD was 1.4, the slightly high SD warrants further investigation at the item level. The Chi Square probability was not significant for any sample (pre, post or stacked) offering further support of the fit of the items to a unidimensional model.

The high *M* and *SD* of the person locations across all samples suggests the items could be better targeted to the persons. The stacked sample has been used for all analysis from this point forward.

Table 1. Summary statistics output for 14 item CMSEI.

	Pre		Post		Stacked	
	ITEMS		ITEMS		ITEMS	
	Location	Fit Residual	Location	Fit Residual	Location	Fit Residual
Mean	0.00	−0.05	0.00	0.06	0.00	0.09
Std. Dev.	2.03	0.86	0.69	0.89	0.98	1.47
	PERSONS		PERSONS		PERSONS	
	Location	Fit Residual	Location	Fit Residual	Location	Fit Residual
Mean	1.22	−0.40	1.95	−0.26	1.59	−0.41
Std. Dev.	1.51	1.30	1.58	1.23	1.72	1.35
	RELIABILITY		RELIABILITY		RELIABILITY	
PSI	0.85		0.88		0.89	
Cronbach's Alpha	0.86		0.89		0.90	

Item and person fit to the model

An individual item analysis revealed item 14 had a high fit residual; however, the Chi Square Probability (0.162) was not significant at the Bonferroni level of adjustment for the 0.05 level. No persons were identified as extreme (having response patterns that do not fit with those predicted by the model).

Violations of independence

In the PCA of the residuals, the principal component loadings showed no relative difference between the first two components, with Eigenvalues of 2.2 and 1.4 respectively. A subtest analysis based on positive and negative loadings of items on the first principal component showed no drop in PSI and the error adjusted correlation between the two analyses was A = 0.98, supporting a unidimensional structure. In the residual correlation matrix, item 10 had a residual correlation above 0.3 with item 12, which may indicate a violation of response independence.

Differential item functioning (DIF)

In analysing the Item Characteristic Curve (ICC) graphs, no items showed potential uniform DIF for programme, age, prior experience working with children, gender or parental status. The ANOVA results supported this interpretation, with an F ratio probability less than 0.001190 being indicative of a statistically significant difference in group means on an item at the 0.05 level.

Response option frequencies

Table 2 summarises the category use of each category: strongly disagree, disagree, agree, and strongly agree; for each item. The category 'strongly disagree' was rarely used and did not meet the recommended minimum of 10 respondents for every item (Linacre 1999).

Table 2. Category frequencies stacked CMSEI.

Item	SD	D	A	SA
1	3	23	96	14
2	4	23	100	11
3	4	19	87	28
4	8	47	68	14
5	3	31	86	18
6	5	53	75	4
7	1	15	87	34
8	4	20	100	14
9	1	29	99	9
10	1	15	85	36
11	1	14	74	48
12	1	14	91	31
13	2	11	73	52
14	5	37	85	10

Category and threshold order

Category thresholds should increase monotonically (Bond and Fox 2015), and this was true for each item in the CMSEI. Inspection of the category probability curves indicated category order for all items, where at least one category was always the most probable category at an incremental value along the *x* axis (Bond and Fox 2015).

The minimum recommended distance for a 4 – point scale of 1.4 logits (Linacre 1999) was met consistently by the Likert scale for all items. For item 10, the distance exceeded the recommended maximum distance of 5 logits between thresholds 1 and 2 and 2 and 3.

Discussion

The 14 item CMSEI was analysed using the responses of 69 pre-service teachers enrolled in the third year of an undergraduate teacher education course. The sample for Rasch analysis consisted of 138 responses, being the combined, matched pre and post responses. The fit statistics suggest the items are measuring a single construct supporting the construct validity of the measure in combination with the face and content validity established during the development phase. Both the PSI (0.89) and Cronbach's alpha (0.90) infer reliability and accuracy of measurement.

The targeting of the items to the persons suggested that many of the items were too easy to endorse for this sample, the high mean person location, lack of items above 1.6 logits and the low usage of the category 'strongly disagree' were indicative of this. There were, however, thresholds for items functioning across the entire range of person abilities. The targeting of the items to the persons may be improved by adding items which are more difficult to endorse. Easy to endorse items in CMSE scales was a theme identified by O'Neill and Stephenson (2011), who suggest adding items which were more challenging to CMSE scales. Figure 4 shows the item difficulty map for the CMSEI scale; with items 4 (There are very few students that I cannot handle), 6 (I can keep defiant students involved in my lessons) and 14 (I am able to explain the rationale, programme components, operation, and evaluation of the behavioural techniques I use) the most difficult items for this sample to positively endorse. Defiant students are a challenge for pre-service teachers (Kher, Lacina-Gifford, and Yandell 2000; Main and Hammond 2008), while the difficultly of item 14 relative to the other items suggests the respondents did not feel that

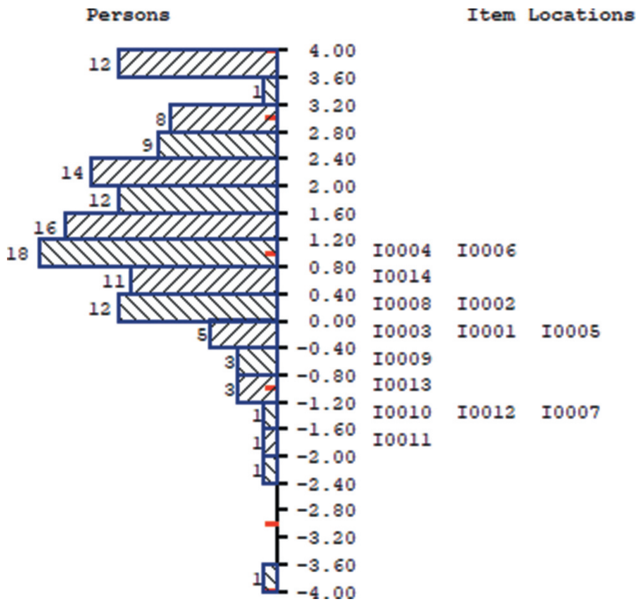

Figure 4. Item map for CMSEI.

they had a good understanding of some aspects of classroom management theory. It may be beneficial to break this item into sub questions relating to rationale, programme components, operation, and evaluation to better inform pre-service curriculum design.

Baker (2005) noted that the in-service teachers in her study were less confident in some of the approaches associated with applied behaviour analysis and functional behaviour assessment, such as documenting student behaviour using formal and systematic methods, using reinforcement hierarchies, and applying reinforcement schedules. To better target both the third and fourth year cohorts, items could be developed around these themes. For example, the use of formal and systematic methods to identify the purpose of behaviour and implementing positive behaviour strategies.

Items should operate independently of one another, and therefore no residual correlations should remain once item difficulty and person ability are accounted for. A possible violation of independence was identified for questions 10 and 12 in the residual correlation between the two items. They were very easy to endorse items for this scale, sitting at location −1.189 and −1.138 logits respectively. Item 10 asked students about their capacity to set appropriate rules while item 12 asked students about their capacity to establish routines. At face value, these items represent different aspects of classroom management, and both items should be retained in the questionnaire to uphold content validity. The removal of either or both items has a negligible effect on the PSI and fit statistics. Cognitive interviewing prior to future dissemination of the questionnaire is required to elicit more understanding regarding how individuals in the target population are interpreting these two questions so that refinements to wording may be made if necessary.

As the instrument was piloted with the third year cohort, the mean of the person estimates at time one (0.94 logits), prior to completing their third year, four week

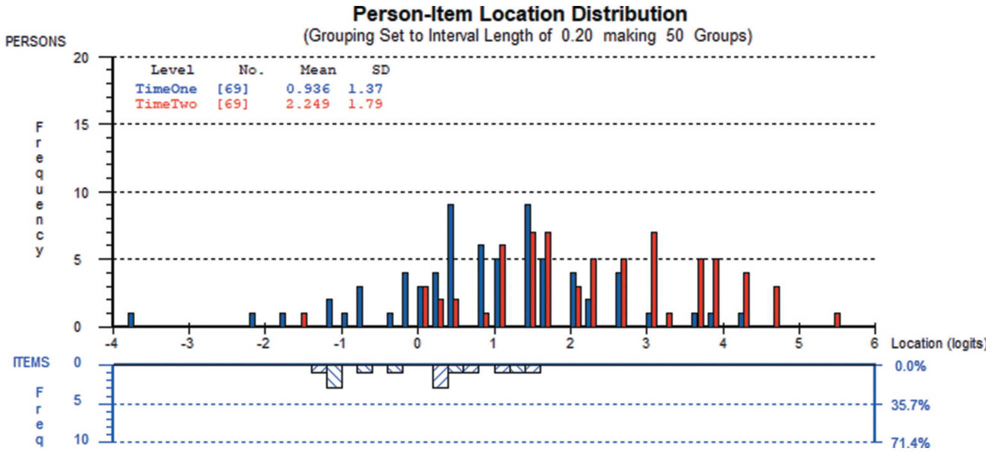

Figure 5. Person-item locations pre and post completion of a third year, four week practicum.

practicum, would theoretically allow for the instrument to be used with a second year cohort without creating a floor effect, and the addition of the suggested difficult to endorse items would protect against a ceiling effect if the questionnaire were implemented with final, fourth year students. Figure 5 shows the person-item locations for the sample at time one and time two, supporting that downward movement would be possible, while more difficult items are required to accurately measure a hypothesised more self-efficacious fourth year cohort. This would allow the students to be tracked across time and for the results to be used in a formative way for pre-service curriculum design, as described by O'Neill and Stephenson (2011).

The Likert scale functioned well for the set of items. Minimum distances between thresholds were consistently met and only item 10 exceeded the upper limit of 5 logits. The low category usage frequencies for 'strongly disagree' are likely due to the sample size and the targeting of the items to the sample, rather than with the selected 4 point scale. This should be resolved with a larger sample (N = 20 persons per item minimum) and the addition of more difficult to endorse items.

Conclusion

This analysis of Main and Hammond's (2008) CMSEI suggests it is a valid and reliable, theory based measure of classroom management self-efficacy for use with pre-service teachers in Australia. The number of items makes it relatively quick to administer and, consequently, it could be used during pre-service coursework activities to provide an indication of how self-efficacy for classroom management is developing. In addition, it can provide an opportunity for pre-service teachers to reflect on their self-efficacy for classroom management. While a valid and reliable measure in its current form, it could be further refined with the addition of items that are more difficult to endorse which would further extend the range of self-efficacy it can accurately measure and allow for the tracking of classroom management self-efficacy across time during pre-service teacher education courses.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ORCID

Eileen V. Slater http://orcid.org/0000-0002-2801-6392

References

Aloe, A. M., L. C. Amo, and M. E. Shanahan. 2014. "Classroom Management Self-Efficacy and Burnout: A Multivariate Meta-analysis." *Educational Psychology Review* 26 (1): 101–126. doi:10.1007/s10648-013-9244-0.

Andrich, D. 1982. "An Index of Person Separation in Latent Trait Theory, the Traditional KR. 20 Index, and the Guttman Scale Response Pattern." *Education Research and Perspectives* 9 (1): 95–104.

Andrich, D., and I. Styles. 2004. "Final report on the psychometric analysis of the Early Development Instrument (EDI) using the Rasch model: A technical paper commissioned for the development of the Australian Early Development Instrument (AEDI)." In. Perth: Murdoch University.

Andrich, D., B. E. Sheridan, and G. Luo. 2014. *RUMM2030-Professional Edition (Rasch Unidimensional Measurement Model, Version 5.4) [Computer Software]*. Perth, Australia: RUMM Laboratory.

Baker, P. H. 2005. "Managing Student Behaviour: How Ready are Teachers to Meet the Challenge?" *American Secondary Education* 33 (3): 51–64. http://www.jstor.org/stable/41064554

Bandura, A. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. 1997. *Self-efficacy: The Exercise of Control*. New York: Freeman.

Bandura, Albert. 2006. "Guide for constructing self-efficacy scales." In *Self-efficacy beliefs of adolescents*, edited by F. Pajares and T. Urdan, 307–37. Greenwich, CT: Information Age Publishing.

Berg, D. A. G., and L. F. Smith. 2016. "Preservice Teacher Self-Efficacy Beliefs." In *Asia-Pacific Perspectives on Teacher Self-Efficacy*, edited by S. Garvis and D. Pendergast, 1–17. Rotterdam: SensePublishers.

Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. In *British Medical Journal, 310*, 170.

Bond, T. G., and C. M. Fox. 2015. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. 3 ed. Hoboken, NJ: Taylor and Francis.

Brophy, J. E. 1976. *Learning from teaching: A developmental perspective*. Boston: Allyn and Bacon

Brophy, J. E. 1996. *Teaching Problem Students*. New York: Guilford.

Brophy, J. E. 2006. "History of research on classroom management." *In Handbook of classroom management: Research, practice, and contemporary issues*, edited by C. M. Evertson and C. S. Weinstein, 17–43. Mahwah, NJ: Lawrence Erlbaum.

Brouwers, A., and W. Tomic. 2000. "A Longitudinal Study of Teacher Burnout and Perceived Self-efficacy in Behaviour Management." *Teaching and Teacher Education* 16: 239–254. doi:10.1016/S0742-051X(99)00057-8.

Brouwers, A., and W. Tomic. 2001. "The Factorial Validity of Scores on the Teacher Interpersonal Self-Efficacy Scale." *Educational and Psychological Measurement* 61 (3): 433–445. doi:10.1177/00131640121971301.

Cano, S. J, L Barrett, J Zajieck, and J. C Hobart. 2011. “Reaching the parts that traditional analysis cannot: Using Rasch to evaluate the DASH.” *Multiple Sclerosis* 17 (2):214–22
Dewey, J. 1899. *The School and Society: Being Three Lectures*. Chicago, IL, US: University of Chicago Press.
Dicke, T., P. D. Parker, H. W. Marsh, M. Kunter, A. Schmeck, and D. Leutner. 2014. “Self-efficacy in Classroom Management, Classroom Disturbances, and Emotional Exhaustion: A Moderated Mediation Analysis of Teacher Candidates.” *Journal of Educational Psychology* 106 (2): 569–583. doi:10.1037/a0035504.
Evertson, C. M., and C. S. Weinstein. 2006. *Handbook of classroom management: Research, practice, and contemporary issues*. Mahwah, NJ: Lawrence Erlbaum.
Evertson, C. M., and C. S. Weinstein, Eds. 2011. *Handbook of Classroom Management: Research, Practice, and Contemporary Issues*. New York: Routledge.
Flower, A., J. W. McKenna, and C. D. Haring. 2017. “Behavior and Classroom Management: Are Teacher Preparation Programs Really Preparing Our Teachers?” *Preventing School Failure: Alternative Education for Children and Youth* 61 (2): 163–169. doi:10.1080/1045988X.2016.1231109.
Goss, P., J. Sonnemann, and K. Griffiths. 2017. *Engaging Students: Creating Classrooms that Improve Learning*. Grattan Institute. https://grattan.edu.au/wp-content/uploads/2017/02/Engaging-students-creating-classrooms-that-improve-learning.pdf
Hagquist, C, M Bruce, and J. P Gustavsson. 2009. “Using the Rasch model in nursing research: An introduction and illustrative example.” *International Journal of Nursing Studies* 46 (3):380–93
Jones, V. F., and L. S. Jones. 2012. *Comprehensive Classroom Management, Creating Communities of Support and Solving Problems*. Upper Saddle River, NJ: Pearson.
Judge, T. A., C. L. Jackson, J. C. Shaw, B. A. Scott, and B. L. Rich. 2007. “Self-efficacy and Work-related Performance: The Integral Role of Individual Differences.” *Journal Of Applied Psychology* 92 (1): 107–127. doi:10.1037/0021-9010.92.1.107.
Kher, N., L. J. Lacina-Gifford, and S. Yandell. 2000. “Preservice Teachers’ Knowledge of Effective Classroom Management Strategies.” In *Paper presented at the American Educational Research Association*, New Orleans, Louisiana. https://eric.ed.gov/?id=ED444941
Kim, K. R., and E. H. Seo. 2018. “The Relationship between Teacher Efficacy and Students’ Academic Achievement: A Meta-analysis.” *Social Behavior and Personality* 46 (4): 529–540. doi:10.2224/sbp.6554.
Korpershoek, H., T. Harms, H. de Boer, M. van Kuijk, and S. Doolaard. 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. doi:10.3102/0034654315626799.
Kounin, J. S. 1970. *Discipline and Group Management in Classrooms*. New York: Holt: Rinehart & Winston.
Kruger, J., and D. Dunning. 1999. “Unskilled and Unaware of It: How Difficulties in Recognizing One’s Own Incompetence Lead to Inflated Self-assessments.” *Journal of Personality and Social Psychology* 77 (6): 1121–1134. doi:10.1037/0022-3514.77.6.1121.
Lewis, T., and G. Sugai. 1999. “Effective Behavior Support: A Systems Approach to Proactive Schoolwide Management.” *Focus on Exceptional Children* 31 (6): 1–24. doi:10.17161/fec.v31i6.6767.
Linacre, J. M. 1999. “Investigating Rating Scale Category Utility.” *Journal of Outcome Measurement* 3 (2): 103–122. https://www.ncbi.nlm.nih.gov/pubmed/10204322
Main, S., and L. Hammond. 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): 28–39. doi:10.14221/ajte.2008v33n4.3.
Martella, R. C., J. R. Nelson, N. E. Marchand-Martella, and M. O`Reilly. 2012. *Comprehensive Behavior Management: School-wide, Classroom, and Individualized Approaches*. 2nd ed. Thousand Oaks, CA: Sage Publications.

Marzano, R. J., J. S. Marzano, and D. J. Pickering. 2003. *Classroom Management that Works: Research-based Strategies for Every Teacher*. Alexandria, VA: Association for Supervision and Curriculum Development.

O'Neill, S. C., and J. Stephenson. 2011. "The Measurement of Classroom Management Self-efficacy: A Review of Measurement Instrument Development and Influences." *Educational Psychology* 31 (3): 261–299. doi:10.1080/01443410.2010.545344.

Page, A., and M. Jones. 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. doi:10.14221/ajte.2018v43n11.5.

Postholm, M. B. 2013. "Classroom Management: What Does Research Tell Us?" *European Educational Research Journal* 12 (3): 389–402. doi:10.2304/eerj.2013.12.3.389.

Rogers, B. 2015. *Classroom Behaviour: A Practical Guide to Effective Teaching, Behaviour Management and Colleague Support*. 4th ed. Los Angeles, CA: SAGE.

Rogers, C. R. 1957. "Personal Thoughts on Teaching and Learning." *Merrill-Palmer Quarterly (1954–-1958)* 3 (4): 241–243. http://www.jstor.org/stable/23080641

Rotter, J. B. 1966. "Generalized Expectancies for Internal versus External Control of Reinforcement." *Psychological Monographs* 80 (1): 1–28. https://www.ncbi.nlm.nih.gov/pubmed/5340840.

Sailor, W., G. Dunlap, G. Sugai, and R. Horner, Eds. 2009. *Handbook of Positive Behavior Supports*. New York: Springer.

Simonsen, B., A. S. MacSuga-Gage, D. E. Briere, J. Freeman, D. Myers, T. M. Scott, and G. Sugai. 2014. "Multitiered Support Framework for Teachers' Classroom-Management Practices: Overviewand Case Study of Building the Triangle for Teachers." *Journal of Positive Behavior Interventions* 16 (3): 179–190. doi:10.1177/1098300713484062.

Skaalvik, E. M., and S. Skaalvik. 2017. "Motivated for Teaching? Associations with School Goal Structure, Teacher Self-efficacy, Job Satisfaction and Emotional Exhaustion." *Teaching and Teacher Education* 67: 152–160. doi:10.1016/j.tate.2017.06.006.

Skinner, B. F. 1953. *Science and Human Behavior*. New York: Free Press.

Subban, P. K., and P. Round. 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. doi:10.14221/ajte.2015v40n5.7.

Talsma, K., B. Schüz, R. Schwarzer, and K. Norris. 2018. "I Believe, Therefore I Achieve (And Vice Versa): A Meta-analytic Cross-lagged Panel Analysis of Self-efficacy and Academic Performance." *Learning and Individual Differences* 61: 136–150. doi:10.1016/j.lindif.2017.11.015.

Tennant, A, and G Conaghan. 2007. "The Rasch measurement model in rheumatology: What is it? Why use it? When should it be applied and what should one look for in a Rasch paper?." *Arthritis & Rheumatism* 57 (8):1358–62

Thorndike, E. L. 1919. *Educational Psychology: Briefer Course*. New York, NY: Columbia University Press.

Tschannen-Moran, M., A. W. Hoy, and W. K. Hoy. 1998. "Teacher Efficacy: Its Meaning and Measure." *Review of Educational Research* 68 (2): 202–248. doi:10.3102/00346543068002202.

Tschannen-Moran, M., and A. Woolfolk-Hoy. 2001. "Teacher Efficacy: Capturing an Elusive Construct." *Teaching and Teacher Education* 17: 783–805. doi:10.1016/S0742-051X(01)00036-1.

Woodcock, S., and A. Reupert. 2012. "A Cross-sectional Study of Student Teachers' Behaviour Management Strategies Throughout Their Training Years." *The Australian Educational Researcher* 39 (2): 159–172. doi:10.1007/s13384-012-0056-x.