Received: 24 July 2023 | Accepted: 11 October 2023

DOI: 10.1111/1467-8578.12494

ORIGINAL ARTICLE

Measuring safety perceptions of students with behavioural problems in special education: A validation study of the safe at school questionnaire

G. H. P. Van der Helm[1,2] | **G.J. Klapwijk**[2] | **J. J. Roest**[1,2] | **C. H. Z. Kuiper**[1,2] | **R. H. J. Scholte**[3] | **G. J. J. M. Stams**[1]

[1]Research Institute for Child Development and Education, University of Amsterdam, Amsterdam, The Netherlands

[2]Kenniscentrum Samen Redzaam, University of Applied Sciences Leiden, Leiden, The Netherlands

[3]Department of Pedagogical Sciences, Radboud University, Nijmegen, The Netherlands

Correspondence
G. H. P. Van der Helm, Research Institute for Child Development and Education, University of Amsterdam, PO Box 15780, 1001 NG Amsterdam, The Netherlands.
Email: helm.vd.p@hsleiden.nl

Abstract

Assessment and monitoring of safety in special education schools is imperative to establish and maintain a safe environment in which students can develop academic and social–emotional skills. The present study describes the development of a student self-report measure, the Safe at School questionnaire. Factorial validity, reliability and concurrent validity was examined in a construction sample of 280 students (68.6% male, age $M = 13.29$, $SD = 2.52$) and a validation sample of 1572 students (77.4% male; age $M = 14.33$, $SD = 2.04$), representing 233 classes from 20 schools for special education in the Netherlands. Multilevel confirmatory factor analysis was used to examine the factor structure of the Safe at School questionnaire. A model with three correlating factors at the within- and between-group level (Rules and Social Norms, Perceived Own Safety and Perceived Safety of Other Students) showed a satisfactory fit to the data. Reliabilities of the scales were good. Also, measurement invariance for gender was demonstrated, indicating no differences in latent means between boys and girls. Implications for practice and future research are discussed.

KEYWORDS

classroom climate, classroom safety, school climate, school safety, special education

Key points

- Monitoring of safety in special education is crucial for maintaining a safe climate and student development.
- Developmental disabilities and trauma warrant the construction and validation of a tailored questionnaire.
- Measuring safety does not result in safe classes, but actively discussing the topic with students does.

INTRODUCTION

In the Netherlands, students with emotional and behavioural problems (EBD) or mild intellectual disabilities (MID) often attend special needs education. These students are characterised not only by severe behavioural problems, but also by adverse childhood experiences (such as parental neglect or domestic violence), psychiatric problems, trauma and a short attention span (Anckarsäter et al., 2007). It is known that children and young people with limited cognitive abilities experience problems with social information processing, which may

result in aggressive behaviour in the classroom (Beld et al., 2019; Van Rest et al., 2019). Also, students attending special education tend to get involved in conflicts in the classroom (such as bullying) as well as other school areas more often than their typically developing peers (Rose et al., 2010).

Exposure to aggressive incidents in the classroom can lead to increased stress and feelings of being unsafe in students and teachers, and puts students at risk of victimisation as well as developing further (aggressive) behaviour problems (Hong & Espelage, 2012; Thomas et al., 2011). Several studies have found that students' feelings of being unsafe at school have a negative impact on their academic achievement, and that feeling safe at school is positively related to student well-being and classroom engagement (Côté-Lussier & Fitzpatrick, 2016; Lacoe, 2016; Nijs et al., 2014). Furthermore, school safety is recognised as an essential dimension of a positive learning climate in schools (Thapa et al., 2013), which is considered a necessary condition for positive academic and social–emotional outcomes of students (Anderson et al., 2004: Beld et al., 2019; Wissink et al., 2014).

Currently, there is a need for measures to routinely monitor safety in schools and in the classroom in order to establish and maintain a safe environment in which students can develop academic and social–emotional skills. In the Netherlands, as of August 2015, measuring safety at school on a yearly basis is compulsory by law. However, measuring students' perceptions of safety in special education is difficult, because of their limited cognitive abilities and short attention span. In addition, existing instruments, such as the California School Climate Survey (Furlong et al., 1991, 2005), comprise many items and use difficult language, rendering them less suitable for use in special education settings. The aim of the present study was to develop a self-report measure to assess student perceptions of safety in the classroom and in school, and examine its psychometric properties in terms of factorial structure, reliability and concurrent validity. This instrument was specifically designed for students with emotional and behavioural disorders and receiving special needs education in the Netherlands.

Classroom safety

Classroom (and school) safety is a much-debated topic (Barrett, 2010; Rom, 1998), especially concerning extreme acts of school violence (such as school shootings), but also suicide attempts by students related to peer victimisation (Phillips, 2007; Schroeder, 2016). However, research on students' perceptions of safety in school remains sparse. Review studies on school safety have mainly focused on school violence (Johnson, 2009; Steffgen et al., 2013). In their review, Thapa et al. (2013) addressed safety in school as an important dimension of school climate. However, many of the studies in their review focused mainly on experiences with school violence or bullying instead of students' perceptions of classroom and school safety in general.

Maslow (cited in Healy, 2016) described feeling socially, emotionally, physically and intellectually safe as a fundamental human need. Lacoe (2016, p. 2) defines students' perceptions of classroom safety as follows: 'how students report feeling inside the classroom, and may include physical or emotional safety and perceptions of a secure environment'. Alternatively, Holley and Steiner (2005) and Thapa et al. (2013) define classroom safety, designated as 'safe space', as an environment in which students are protected from psychological and emotional harm. From this perspective, preventing violence is an important focus of establishing safe spaces in schools. Rom (1998) distinguishes four elements of a 'safe space'. First, a 'safe place' refers to a group of children; second, the physical space of the classroom stands for social connectedness; third, the safe space is characterised as comfortable; and fourth, students in a safe space are able to do better work and learn more. Additionally, Thapa et al. (2013) state that creating clear boundaries by setting rules and norms, which are perceived as fair by students, supports a positive school climate and perceived safety. Oberle et al. (2016) also claim that safety is a basic condition that enables the development of a positive learning climate in schools, since a safe environment serves as a foundation for creating a positive learning environment, allowing students and teachers to learn from each other.

Students' feelings of safety are shaped by experiences in the immediate school and classroom context, such as behavioural incidents, procedural fairness, and enforcement of classroom rules and norms by the teacher (Lacoe, 2016). According to self-determination theory, feelings of safety are also a result of students' interpretations of social cues (connectedness), feelings of social competence, self-worth and autonomy (Ryan & Deci, 2017). Emmer and Sabornie (2014) suggest that students are an important part of classroom management in special educational settings, which requires specific attention by teachers to prevent challenging behaviour by children, and to intervene when problems occur. Ryan et al. (2001) also state that the teacher has an important task in establishing rules and norms in the classroom. Furthermore, Emmer and Sabornie (2014) explain that, in special education classrooms, it is important for teachers to explicitly teach children rules on how to be respectful and to explicate safety rules. In addition, Marzano and Gaddy (2005) state that establishing classroom rules and norms is necessary for effective classroom management, supporting predictability and students' feelings of safety. Therefore, safety at school and in the classroom can be defined as a multifaceted construct comprising students' own feelings of safety, a safe social atmosphere in the classroom, and clear structure and rules (Bosworth et al., 2011; Marzano & Gaddy, 2005).

Measuring student perceptions of safety in special education

Students with MID and/or EBD often have a limited attention span, difficulty prioritising information and weighing its importance, an impaired verbal- and long-term working memory, and limited meta-cognitive abilities, and tend to be less able to see things from the perspective of another person (Kleinert et al., 2009; Willner et al., 2010). Van Rest et al. (2019) and Beld et al. (2019) found that young people with EBD tend to process social information or social cues in a more hostile way compared to typically developing peers. Therefore, young people with EBD are at risk of developing persistent aggressive behaviour, which can lead to rule breaking and disruptive behaviour in the classroom. Repeated failure to engage in prosocial interactions with peers and teachers in the classroom can lead to feelings of incompetence, low self-esteem, and an aggressive or submissive problem-solving style (Van der Helm et al., 2011). For these young people, establishing a classroom and school environment in which students feel physically and emotionally safe, by providing clear rules and social norms that are perceived as fair by students, is important for the development of social skills and improvement of executive functioning (Brock et al., 2018).

Considering the characteristics of students in special education, a self-report questionnaire for use in this population should be short in order not to overstretch attention span, and wording should be simple in the 'here and now' formulation and not draw on things that happened some time ago. Also, the items should not draw upon meta-cognitive thinking, because these students have trouble ordering, ranking and differentiating information. Existing student self-report instruments often consist of many items, which makes them impractical for students who have a short attention span. Also, these measures often include meta-cognitive constructs, such as asking students how often certain types of problems have occurred in school, remembering things that happened a month ago, what happened 'more often', or what ideas and suggestions students have to improve school safety (Furlong et al., 2005; Hurford et al., 2010).

Student self-report measures of school safety can be used in practice to generalise findings to the school or classroom level, because the perceptions of classroom and school safety vary across students, and the perceptions of classroom and school safety of students from the same class are expected to be more similar than the perceptions of students from different classes. However, regular factor analytic techniques applied in psychometric studies on student self-report measures often do not take into account the nested nature of the data (e.g. students within classrooms). In regular (single-level) factor analysis, observations (students) are treated as independent (from different classrooms), disregarding the multilevel structure. In recent years, multilevel techniques have become available, which can be applied to factor analysis (Byrne, 2012; Hox, 2002). In multilevel factor analysis, the factor structure of a measure can be examined at both the within- (individual) and between-group (classroom) level (Huang, 2017). Given the fact that scales to assess school and classroom safety are used to measure safety at the classroom level, multilevel factor analysis is needed to examine the factor structure of the scales at the between-group level in addition to the factor structure at the within-group level.

Present study

In special educational settings in particular, it is important to monitor students' safety perceptions, since teachers and students are more often faced with disruptive behaviour and aggressive incidents in the classroom, which can lead to students and teachers feeling unsafe. The aim of this study was to develop a self-report measure to assess student perceptions of safety in the classroom and in school, specifically designed for students with behavioural problems and MID receiving special education. Student-perceived classroom safety was operationalised in terms of students' feelings of physical and emotional safety and unsafety inside the classroom and school, perceptions of clear rules and norms, and perceptions of fairness of rules.

The factor structure and reliability of the measure was examined as well as the concurrent validity, through examining the relation between the scales and aspects of classroom climate. First, a construction sample was used to examine the factor structure and reliability of the factors. Second, using a validation sample, the factor structure found in the construction sample was replicated. The larger sample size of the validation sample allowed us to use multilevel confirmatory factor analysis to examine the factor structure of the factors at the within- and between-group (i.e. classroom) level. Additionally, multiple group confirmatory factor analyses were conducted using the validation sample to test several levels of measurement invariance between boys and girls, since differences in interpretations of the items may occur because of gender-specific responses to social stress and feelings of unsafety, that is, a fight-or-flight response in boys versus a tend-and-befriend response in girls (Cardoso et al., 2013; Taylor, 2006), or differences in social information processing and social cognition in general (Adrian et al., 2010).

METHOD

Participants

The construction sample consisted of $N = 280$ students, $n = 192$ (68.6%) boys, $n = 88$ (31.4%) girls; the mean age of respondents was 13.29 years ($SD = 2.52$, $Min = 8$,

Max = 17), and 88.6% of the students were of Dutch nationality. The sample included young people from four schools for special education in the Netherlands. These students were referred to special education because of EBD and/or an intellectual disability. Students represented 38 classes with an average cluster size of 7 (range = 1–13 students per class).

The validation sample consisted of $N = 1572$ students, $n = 1217$ (77.4%) boys, $n = 355$ (22.6%) girls; mean age of respondents was 14.33 years ($SD = 2.04$, *Min* = 8, *Max* = 21); 91.0% of the students were of Dutch nationality. The sample included young people from 20 schools for special education in the Netherlands. Students represented 233 classes with an average cluster size of 6.7 (range = 1–15 students per class).

Procedure

Data were obtained from October 2015 until May 2017 for the construction sample, and from September 2017 until January 2019 for the validation sample. Participants filled out a questionnaire to measure perceived safety in school and in the classroom, as well as classroom climate. All students participated voluntarily and signed an informed consent form. Questionnaires were assigned a number in order to guarantee anonymity of the participants. Participants were told their answers would only be accessed by the researchers, and were assured that teachers and other staff members would not have access to their answers. Participants did not receive any compensation for their participation in the study.

Measures

The safe at school questionnaire

A self-report questionnaire was developed to assess students' perception of safety in the classroom and in school. An item pool was generated with items referring to students' own feelings of safety and unsafety in the school (for example, 'I feel safe in the school surroundings', 'I am afraid at school'), and items referring to perceptions of rules and social norms in the classroom (for example, 'The rules in the classroom are clear', 'The rules in the classroom are fair' and 'We discuss in the classroom how we behave towards each other').

Initially, only items referring to students' own safety were formulated. However, when a draft of the questionnaire was pilot-tested in a school that was known to have serious issues regarding school safety, students reported that they felt safe themselves, because they could defend themselves against an attack. We therefore developed additional items, asking whether students thought other students felt safe or were threatened in the classroom, in the school hallway and in the school surroundings.

Items were rated on a five-point Likert-type scale, ranging from 1 ('I do not agree') to 5 ('I totally agree'). In addition to filling out the questionnaire, students were also asked to evaluate school safety by giving single item ratings between 1 (very poor) and 10 (excellent).

Classroom climate

The Special Education Classroom Climate Inventory-Revised (SECCI-R; Beld et al., 2019) was used to measure student perceptions of classroom climate. The SECCI-R consists of 21 items representing five scales. The Teacher Support Scale (five items) assesses the responsiveness of teachers to specific needs of students. Negative Student Interactions (five items) assesses negative interactions among students. Positive Student Affiliation (four items) assesses positive student affiliations in the classroom. Structured Classroom Environment (three items) assesses the degree to which basic conditions are created that enable students to focus their attention on learning. Growth (four items) assesses the degree to which students believe they are learning and are able to develop in the classroom. All items are rated on a five-point Likert-type scale, ranging from 1 ('I do not agree') to 5 ('I totally agree'). Higher scores are indicative of more teacher support, growth, positive student affiliation, structured classroom environment and negative student interactions. One item on the Growth scale was reverse-scored, meaning that a higher score represents a negative outcome.

Statistical analyses

Using the construction sample, construct validity was examined by means of confirmatory factor analysis (CFA), using Mplus software, version 6.12 (Muthén & Muthén, 1998–2012). To account for non-normally distributed ordinal variables, the mean and variance-adjusted weighted least squares (WLSMV) estimation procedure was chosen (Li, 2015). Pairwise deletion was used to handle missing data. Modification indices, giving the expected drop in chi-square if the parameter in question is freely estimated, were used to improve model fit. Residual variances of items that were similarly worded were allowed to correlate.

In the validation sample, multilevel confirmatory factor analysis (MCFA) was conducted. First, the between-group and within-group covariance matrices were computed using full information maximum likelihood (FIML) estimation. Then, a single-level CFA was conducted using the within-group covariance matrix to examine the hypothesised three-factor model. Also, the between-group covariance matrix was used to test the three factor-model. Next, a multilevel model was specified, in which the within- and between-level part

of the model were examined simultaneously. Several multilevel models were specified, such as a three-factor model at each level, and a three-factor model at level-1 and fewer factors at level-2. The rationale for looking at different models was that previous studies using MCFA have found a smaller number of level-2 factors relative to level-1 (Hox, 2002). Four items of the factors Perceived Own Safety (two items) and Perceived Safety of Other Students (two items) were recoded, such that for all items a higher score indicated a positive outcome. Research on MCFA has found that reverse-scored items may cause convergence problems (Gustafson & Stahl, 2005). In MCFA models, negative residual variance at level-2 is a common problem, which can result in non-convergence of the model (Kim et al., 2016). The final MCFA model resulted in negative residual variance at level-2 for four items (items 1, 12, 16 and 17). Because the residual variance of these items at level-2 was close to zero and non-significant, the residual variance of these items was fixed to zero, which is a recommended practice when using MCFA (Hox, 2002).

Model fit was evaluated by various model fit indices; comparative fit index (CFI), Tucker–Lewis index (TLI) and the root mean square residual (RMSEA). For the MCFA models, standardised root mean square residual (SRMR) at the within-group level ($SRMR_W$) and between-group level ($SRMR_B$) were examined. Cut-off values of CFI>0.95, TLI>0.95, RMSEA <0.05 and SRMR <0.08 are required for good model fit, and values of CFI>0.90, TLI>0.90 and RMSEA <0.08 are indicative of acceptable model fit (Hu & Bentler, 1999; Kline, 2015).

Reliability analyses (Cronbach's alpha) were conducted using the 'mcfa.input' function and syntax (Huang, 2017) in the R environment (version 3.4.1; R Core Team, 2017). This function generates the within- and between-group covariance matrices, and allows computation of Cronbach's alpha for level-1 factors (using the within-group covariance matrix) and level-2 factors (using the adjusted between-group covariance matrix).

Intraclass correlations (ICC-1) of the scales were computed, to examine the amount of variability between groups (i.e. between classes), and the degree of non-independence of the data (Raudenbush & Bryk, 2002). The ICC-1 value refers to the proportion of variance explained by the group. Values greater than zero are indicative of nested data structures, in which case multilevel analysis is warranted (Byrne, 2012). Also, ICC-2 was calculated for the factor scores to examine inter-rater reliability of scores between students within classes (Bliese, 2000). ICC-2 (inter-rater reliability) refers to the extent to which an individual score can be considered a reliable assessment of a group-level construct. Also, within-group rater agreement for multi-item measures (rWG(J); James et al., 1984) was computed to examine the agreement in scores across group members on all items of a scale. ICC-2 and rWG(J)≥0.70 are generally accepted as good (Bliese, 2000). ICC-1, ICC-2 and rWG(J) were calculated using the *bruceR* package (Bao, 2023).

Convergent validity was examined by calculating Pearson's *r* correlations between school safety scales and the report mark (between 1 and 10) for perceived school safety. A positive moderate to strong correlation between the Rules and Social Norms, Perceived Own Safety and Perceived Safety of Other Students scales, and the report mark would be indicative of convergent validity of the three scales.

To examine concurrent validity, correlations between the safety scales and the SECCI-R scales were examined. Concurrent validity is demonstrated when perceived school safety positively correlates with teacher support, growth, positive student affiliation, structured classroom environment, and the report mark for perceived safety. Also, concurrent validity is demonstrated when perceived school safety correlates inversely with disruptive behaviour. Pearson's correlations of $r=0.10–0.30$ are seen as small, $r=0.30–0.50$ are seen as moderate, and $r>0.50$ are seen as large (Cohen, 1988).

Multiple group CFA was conducted for groups based on gender to examine differences in latent means between boys and girls, for which scalar invariance is required (Van de Schoot et al., 2012). We followed the procedures for testing measurement invariance with ordinal variables as proposed by Wu and Estabrook (2016), and followed the steps and syntax as outlined by Svetina et al. (2020), using the *lavaan* (Rosseel, 2012) and *semTools* (Jorgensen et al., 2018) packages in the R environment (R Core Team, 2017). First, two separate models were fitted for each group. Then, an unconstrained multiple group model was fitted to examine configural invariance. In this model, the factor variances and scales were fixed to 1, the factor means and intercepts were fixed to 0 in each group, and all item loadings and thresholds (four per item) were freely estimated. In the second step, a model was fitted with thresholds constrained to be equal across groups (threshold invariance). The factor variances were fixed to 1 and factor means were fixed to 0 in both groups. Factor scales were fixed to 1 and intercept means were fixed to 0 in the reference group (boys), and freely estimated in the other group (girls). Third, a model was fitted with loadings and thresholds constrained to be equal across groups. In the reference group, the factor variances, scales and intercept means were fixed to 1, but freely estimated in the other group. Factor means were fixed to 0 in both groups.

To examine change in model fit of multiple group models, the lavTestLRT() function in *lavaan* was used, which tests the change in chi-square when models are nested. A non-significant chi-square test is indicative of negligible change in model fit across models, which demonstrates measurement invariance between groups. Because chi-square is sensitive to sample size, change in CFI (Δ CFI) was also used to examine change in model fit of multiple group analyses. A change in CFI by 0.01

or more is indicative of non-invariance between groups (Cheung & Rensvold, 2002).

RESULTS

Missing data and outliers

In the construction sample, missing values for the individual items ranged between zero and 20 (7.1%). Little's MCAR test ($\chi^2_{260}=289.52$, $p=0.101$) indicated that missing data were completely at random. In the validation sample, missing values for the individual items ranged between 16 (1.0%) and 48 (2.9%). Little's MCAR test ($\chi^2_{786}=982.78$, $p<0.001$) was significant, indicating that missing values were not missing completely at random.

Skewness and kurtosis values in both samples indicated that several items were positively skewed (item 7 and item 8) with kurtosis values >3.0, indicating that these items were non-normally distributed. ICCs ranged between 0.00 and 0.33 in the construction sample, and between 0.08 and 0.31 in the validation sample.

Confirmatory factor analysis: Construction sample

First, a single level CFA was conducted using the total covariance matrix of the construction sample. A three-factor model was specified, consisting of Rules and Social Norms, Perceived Own Safety and Perceived Safety of Other Students. Initial fit of the model was acceptable: $\chi^2_{116}=356.78$; $p<0.001$, CFI=0.923, TLI=0.910, RMSEA=0.086 (90% CI=0.076, 0.097). Model fit was improved by correlating residual variance of similarly worded items (for example, 'I feel safe in school' and 'I feel safe in the hallway'), based on relatively large modification indices (>40). These modifications of the model resulted in a better model fit: $\chi^2_{113}=288.25$; $p<0.001$, CFI=0.944, TLI=0.933, RMSEA=0.075 (90% CI=0.064, 0.085).

Internal consistency (Cronbach's alpha) of the factors was good: Perceived Own Safety $\alpha=0.75$; Perceived Safety of Other Students $\alpha=0.70$; and Rules and Social Norms $\alpha=0.89$. The factors were significantly and moderately correlated: Perceived Own Safety and Perceived Safety of Other Students ($r=0.43$), Perceived Safety of Other Students and Rules and Social Norms ($r=0.42$), and Perceived Own Safety and Rules and Social Norms ($r=0.44$).

A relatively high mean score was found for Perceived Own Safety ($M=4.36$, $SD=0.82$) and lower mean scores for Perceived Safety of Other Students ($M=3.63$, $SD=0.87$) and Rules and Social Norms ($M=3.80$, $SD=0.88$). The scales demonstrated relatively large ICC-1 values (between 0.13 and 0.34), indicating that a substantial percentage of variance in scores could be attributed to the classroom level. Further, ICC-2 and rWG(J) values were calculated. ICC-2, referring to the reliability of scores as a group construct, indicated good reliability for Rules and Social Norms. Reliability for Perceived Own Safety and Perceived Safety of Other Students was generally poor (values <0.60). rWG(J) values indicated moderate to good within-group agreement on items of the scales with values>0.65. Descriptive statistics of the scales and correlations between scales are depicted in Table 1.

Multilevel confirmatory factor analysis: Validation sample

Next, using the validation sample, we conducted multilevel confirmatory factor analyses (for all MCFA models, see Table 2). The model found in the construction sample was replicated, which resulted in acceptable model fit: $\chi^2_{113}=1136.17$; $p<0.001$, CFI=0.957, TLI=0.947, RMSEA=0.077 (90% CI=0.073, 0.081). Next, we examined the three-factor model at the between-group level, however, this resulted in poor model fit. Also, alternative models such as a one or two-factor model resulted in poor model fit. Next, a multilevel factor model was specified with a three-factor model at the within-group level and three factors at the between-group level, which showed an acceptable overall fit to the data: $\chi^2_{229}=1140.14$; $p<0.001$, CFI=0.934, TLI=0.921, RMSEA=0.050, $SRMR_W=0.060$, $SRMR_B=0.134$. However, these results also suggest that the between-group part of the model demonstrated poor model fit. The factors at the within-group level were significantly and moderately correlated: Perceived Own Safety and Perceived Safety of Other Students ($r=0.39$), Perceived Safety of Other Students and Rules and Social Norms ($r=0.42$), and Perceived Own Safety and Rules and Social Norms ($r=0.39$). At the between-group level, factors were also positively and significantly correlated: Perceived Own Safety and Perceived Safety of Other Students ($r=0.94$), Perceived Safety of Other Students and Rules and Social Norms ($r=0.70$), and Perceived Own Safety and Rules and Social Norms ($r=0.79$). Reliability was good, with Cronbach's alpha at the within-level for Rules and Social Norms $\alpha=0.85$, Perceived Own Safety $\alpha=0.76$ and Perceived Safety of Other Students $\alpha=0.74$, and between-level alphas of 0.97, 0.84 and 0.97, respectively.

Mean scale scores were similar to the values found in the construction sample; Perceived Own Safety: $M=4.39$, $SD=0.77$; Perceived Safety of Other Students: $M=3.65$, $SD=0.91$; Rules and Social Norms: $M=3.80$, $SD=0.79$. ICC-1 values ranged between 0.08 and 0.33; ICC-2 values indicated good reliability for Rules and Social Norms and poor reliability for Perceived Own Safety and Perceived Safety of Other Students (values <0.50). rWG(J) values indicated moderate-to-good within-group agreement on items of the scales with values

TABLE 1 Descriptive statistics and correlations for study variables (construction sample).

	Descriptives						Correlations							
Scale	***n***	***M***	***SD***	***ICC-1***	***ICC-2***	***rWG(J)***	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
1. Rules and social norms	272	3.63	0.87	0.34	0.74	0.79	0.44**	0.42**	0.76**	0.55**	0.68**	0.52**	−0.49**	0.57**
2. Perceived own safety	275	4.36	0.82	0.17	0.55	0.69		0.43**	0.34**	0.30**	0.32**	0.50**	−0.38**	0.49**
3. Perceived safety of other students	258	3.80	0.87	0.13	0.47	0.65			0.26**	0.26**	0.31**	0.28**	−0.40**	0.42**
4. Teacher support	277	4.12	0.99	0.34	0.75	0.76				0.53**	0.55**	0.37**	−0.22**	0.43**
5. Growth	272	3.35	0.80	0.14	0.51	0.50					0.33**	0.27**	−0.04	0.40**
6. Structured classroom environment	275	3.47	0.95	0.20	0.60	0.65						0.49**	−0.44**	0.37**
7. Positive student affiliation	279	3.94	0.94	0.21	0.62	0.73							−0.48**	0.37**
8. Negative student interactions	280	2.79	1.04	0.46	0.82	0.66								−0.41**
9. Report mark school safety	270	7.68	2.39	0.16	0.54	–								

Abbreviations: ICC, intraclass correlation coefficient; rWG(J), within-group rater agreement for multi-item measures.
**$p<0.01$.

TABLE 2 Fit measures of multilevel confirmatory factor analysis models.

Multilevel CFA models	χ^2(df)*	CFI	TLI	RMSEA	$SRMR_W$	$SRMR_B$
3 factors within /1 factor between	1123.46 (232)	0.935	0.924	0.050	0.060	0.169
3 factors within /2 factors between	1122.47 (231)	0.935	0.923	0.050	0.060	0.136
3 factors within /3 factors between	1140.14 (229)	0.934	0.921	0.050	0.060	0.134

Abbreviations: χ^2, chi-square; CFA, confirmatory factor analysis; CFI, comparative fit index; df, degrees of freedom; RMSEA, root mean square error of approximation; SRMR, standardised root mean square residual; $SRMR_B$, SRMR between-group; $SRMR_W$, SRMR within-group; TLI, Tucker–Lewis Index.
*All chi-square values are statistically significant at $p<0.001$.

>0.63. Descriptive statistics of the scales and correlations between scales are depicted in Table 3.

Concurrent and convergent validity

Concurrent validity was examined by calculating Pearson's *r* correlations between the safety scales and classroom climate in both the construction and validation sample (Tables 1 and 3). Results indicated that Rules and Social Norms was positively and moderately correlated with Positive Student Affiliation and Growth, and strongly correlated with Teacher Support and Structured Classroom Environment. Small but significant and positive correlations were found between Perceived Own Safety, Perceived Safety of Other Students, and Positive Student Affiliation, Growth, Teacher Support and Structured Classroom Environment. Further small but significant negative correlations were found between Disruptive Behaviour and Rules and Social Norms, Perceived Own Safety and Perceived Safety of Other Students.

Convergent validity was examined by calculating Pearson's *r* correlations between the safety scales and the report mark for perceived school safety. Results indicated that Rules and Social Norms, Perceived Own Safety and Perceived Safety of Other Students were positively and moderately correlated with the report mark for perceived school safety.

Measurement invariance

Next, several multiple group CFA models were examined to test for measurement invariance between boys and girls (see Table 4 for fit indices of all specified models). In the first step, two separate CFAs were conducted for boys ($\chi^2_{113}=1171.81$; $p<0.001$, CFI = 0.960; TLI = 0.952; RMSEA = 0.079 [90% CI = 0.074, 0.084]) and girls ($\chi^2_{113}=483.47$; $p<0.001$, CFI = 0.951; TLI = 0.941; RMSEA = 0.090 [90% CI = 0.081, 0.100]), indicating a good model fit for both groups. Also, a model without constraints across groups displayed a good fit (Model 1: configural invariance). Model 2, representing thresholds (weak) invariance, demonstrated a good fit, and did not fit significantly worse than the configural invariance model. Model 3, representing thresholds and loadings (strong) invariance, also did not fit significantly worse than the weak invariance model. Because strong invariance allows meaningful comparison of latent factor means, we tested for differences in mean factor scores between groups. No significant differences were found in mean factor scores between boys and girls.

TABLE 3 Descriptive statistics and correlations for study variables (validation sample).

	Descriptives						Correlations							
Scale	***n***	***M***	***SD***	***ICC-1***	***ICC-2***	**rWG(J)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
1. Rules and social norms	1492	3.80	0.79	0.33	0.73	0.85	0.39**	0.42**	0.73**	0.48**	0.68**	0.50**	−0.39**	0.50**
2. Perceived own safety	1522	4.39	0.77	0.08	0.35	0.72		0.39**	0.26**	0.14**	0.31**	0.45**	−0.40**	0.49**
3. Perceived safety of other students	1489	3.65	0.91	0.13	0.47	0.63			0.28**	0.21**	0.34**	0.24**	−0.37**	0.44**
4. Teacher support	1540	4.25	0.87	0.33	0.73	0.81				0.47**	0.53**	0.37**	−0.19**	0.38**
5. Growth	1525	3.40	0.68	0.14	0.50	0.61					0.40**	0.28**	−0.06*	0.31**
6. Structured classroom environment	1536	3.55	0.93	0.20	0.60	0.66						0.41**	−0.38**	0.39**
7. Positive student affiliation	1538	3.96	0.90	0.13	0.47	0.72							−0.34**	0.38**
8. Negative student interactions	1563	2.74	0.97	0.34	0.75	0.67								−0.33**
9. Report mark school safety	1535	7.66	2.17	0.10	0.41	–								

Abbreviations: ICC, intraclass correlation coefficient; rWG(J), within-group rater agreement for multi-item measures.
*$p<0.05$; **$p<0.01$.

TABLE 4 Fit measures of multiple group confirmatory factor analysis models for gender.

Tests for measurement invariance	χ^2**(df)***	**CFI**	**TLI**	**RMSEA**	$\Delta\chi^2$**(df)**	***p***	**ΔCFI**	**ΔRMSEA**
Model 1: configural invariance	1273.97 (226)	0.961	0.953	0.093	–	–	–	–
Model 2: threshold invariance	1280.18 (260)	0.960	0.958	0.088	24.62 (34)	0.881	−0.001	−0.005
Model 3: threshold and loadings invariance	1293.60 (274)	0.963	0.964	0.081	11.81 (14)	0.622	+0.003	−0.007

Abbreviations: χ^2, chi-square; CFI, comparative fit index; df, degrees of freedom; RMSEA, root mean square error of approximation; TLI, Tucker–Lewis Index.
*All chi-square values are statistically significant at $p<0.001$.

DISCUSSION

In a two-step validation process, the factor structure, reliability and concurrent validity of the Safe at School (SAS) questionnaire for special education was examined. Using multilevel confirmatory multifactorial analysis, a three-factor model (Rules and Social Norms, Perceived Own Safety and Perceived Safety of Other Students) yielded a satisfactory fit to the data regarding the within-group (individual) level. However, model fit at the between-group level was unsatisfactory. Also, the scales demonstrated good reliability, and evidence was found for concurrent and convergent validity. Measurement invariance across gender was demonstrated, indicating no differences in latent factor means between boys and girls.

There are several limitations to this study that need to be acknowledged. First, although the main aim of the study was to assess construct validity and reliability of the SAS for students with MID and EBD, student and school characteristics may be differentially related to student-perceived school safety. Future studies should examine possible differences in perceived safety between different subgroups, addressing within-group (e.g. gender, age, IQ, diagnosis) and between-group (school size, school type, etc.) variables. Second, it cannot be ruled out that the participants did not understand some of the questions. Neither can the possibility of socially desirable answers be excluded. However, consistency in answering patterns, the fact that the questionnaire contains both positively and negatively formulated items, and concurrent and convergent validity suggest that the influence of social desirability was minimised. Future studies should assess convergent validity of the SAS with a qualitative assessment of students' perception of safety, such as interviews. Concurrent validity can be further assessed by examining the relation between student perceptions of safety and the occurrence of aggressive incidents or bullying in the classroom. Predictive validity can also be established by examining the relation between safety and academic achievement, as studies have found that feelings of safety are related to better school results (Lacoe, 2016).

Despite these limitations, this short and easy-to-comprehend questionnaire is possibly not only suited for special education, but also for measuring perceived safety in mainstream schools. The SAS can also be used together with a measure of classroom climate in practice-orientated research to monitor and improve safety and classroom climate in schools. Routinely monitoring school safety and classroom climate has been found to improve the academic and social development of students (Beld et al., 2019; Maras et al., 2012). Discussing how students perceive aspects of the classroom climate and safety in the classroom and in school, with a focus on sharing experiences and formulating recommendations

for improvements, can be a powerful tool in establishing a positive and safe classroom and school climate (Beld et al., 2019). For instance, when teachers or school psychologists have concerns regarding norm-transgressive behaviour in the classroom or bullying in school, the results of the SAS questionnaire may be used as input for a class discussion of how students experience the rules and norms, and how to improve social interactions among students. Also, actively discussing classroom and school safety may improve feelings of mutual trust and acceptance, and reduce weapon-carrying at school, as opposed to physical security through scanning devices which, according to scarce research, can increase fear (Blosnich & Bossarte, 2011, 2012; Finn & Servoss, 2014).

The present study found preliminary evidence for the internal structure and reliability of the SAS in a group of students with MID and/or EBD attending schools for special education. It is assumed that repeated measurement of school and classroom safety by means of the SAS, discussing the outcomes with teachers and students, can contribute to the improvement of school safety. It requires effort from teachers and schools to establish and maintain a safe school environment for children with MID or EBD, but such an effort is essential in order to stimulate the emotional, social and cognitive development of students in special education.

ACKNOWLEDGEMENTS

We thank all schools of De Onderwijsspecialisten, Aloysius and Horizon for their cooperation.

FUNDING INFORMATION

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

CONFLICT OF INTEREST STATEMENT

All authors declare that they endorse and have acted upon the Netherlands Code of Conduct for Research Integrity (2018) and the European Code of Conduct for Research Integrity (2017), and declare that they have not been influenced by potential conflicts of interest with respect to the authorship or publication of this article.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

The research was approved by Leiden University of Applied Sciences Ethical Commission.

ORCID

G. H. P. Van der Helm https://orcid.org/0000-0001-5425-8478

REFERENCES

Adrian, M., Lyon, A.R., Oti, R. & Tininenko, J. (2010) Developmental foundations and clinical applications of social information processing: a review. *Marriage and Family Review*, 46(5), 327–345. Available from: https://doi.org/10.1080/01494929.2010.527809

Anckarsäter, H., Nilsson, T., Ståhlberg, O., Gustafson, M., Saury, M.J., Råstam, M. et al. (2007) Prevalences and configurations of mental disorders among institutionalized adolescents. *Developmental Neurorehabilitation*, 10(1), 57–65. Available from: https://doi.org/10.1080/13638490600864157

Anderson, A., Hamilton, R.J. & Hattie, J. (2004) Classroom climate and motivated behaviour in secondary schools. *Learning Environments Research*, 7, 211–225. Available from: https://doi.org/10.1007/s10984-004-3292-9

Bao, H.W.S. (2023) *bruceR: broadly useful convenient and efficient R functions*. R package version 0.8.x.

Barrett, B.J. (2010) Is 'safety' dangerous? A critical examination of the classroom as safe space. *Canadian Journal for the Scholarship of Teaching and Learning*, 1(1), 9. Available from: https://doi.org/10.5206/cjsotl-rcacea.2010.1.9

Beld, M.H.M., Kuiper, C.H.Z., Van Der Helm, G.H.P., De Swart, J.J.W., Stams, G.J.J.M. & Roest, J.J. (2019) Classroom climate, identification with school, and general self-worth predict academic self-concept in students attending residential schools for special education. *Residential Treatment for Children and Youth*, 38(5), 1–16. Available from: https://doi.org/10.1080/0886571X.2019.1696262

Bliese, P.D. (2000) Within-group agreement, non-independence, and reliability: implications for data aggregation and analysis. In: Klein, K.J. & Kozlowski, S.W. (Eds.) *Multilevel theory, research, and methods in organizations*. San Francisco, CA: Jossey-Bass, pp. 349–381.

Blosnich, J. & Bossarte, R. (2011) Low-level violence in schools: is there an association between school safety measures and peer victimization? *Journal of School Health*, 81(2), 107–113. Available from: https://doi.org/10.1111/j.1746-1561.2010.00567.x

Blosnich, J. & Bossarte, R. (2012) Drivers of disparity: differences in socially based risk factors of self-injurious and suicidal behaviors among sexual minority college students. *Journal of American College Health*, 60(2), 141–149. Available from: https://doi.org/10.1080/07448481.2011.623332

Bosworth, K., Ford, L. & Hernandez, D. (2011) School climate factors contributing to student and faculty perceptions of safety and select Arizona schools. *Journal of School Health*, 1(4), 194–201. Available from: https://doi.org/10.1111/j.1746-1561.2010.00579.x

Brock, L.L., Murrah, W.M., Cottone, E.A., Mashburn, A.J. & Grissmer, D.W. (2018) An after-school intervention targeting executive function and visuospatial skills also improves classroom behavior. *International Journal of Behavioral Development*, 42(5), 474–484. Available from: https://doi.org/10.1177/0165025417738057

Byrne, B. (2012) *Structural equation modeling with Mplus: basic concepts, applications, and programming*. New York: Routledge.

Cardoso, C., Ellenbogen, M.A., Serravalle, L. & Linnen, A.-M. (2013) Stress-induced negative mood moderates the relation between oxytocin administration and trust: evidence for the tend-and-befriend response to stress? *Psychoneuroendocrinology*, 38(11), 2800–2804. Available from: https://doi.org/10.1016/j.psyneuen.2013.05.006

Cheung, G.W. & Rensvold, R.B. (2002) Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233–255. Available from: https://doi.org/10.1207/S15328007SEM0902_5

Cohen, J. (1988) *Statistical power analysis for the behavioral sciences*, 2nd edition. Mahwah: Erlbaum.

Côté-Lussier, C. & Fitzpatrick, C. (2016) Feelings of safety at school, socioemotional functioning, and classroom engagement. *Journal of Adolescent Health*, 58(5), 543–550. Available from: https://doi.org/10.1016/j.jadohealth.2016.01.003

Emmer, E. & Sabornie, E.J. (2014) *Handbook of classroom management*. New York: Routledge.

Finn, J.D. & Servoss, T.J. (2014) Misbehavior, suspensions, and security measures in high school: racial/ethnic and gender differences. *Journal of Applied Research on Children*, 5(2), 11. Available from: https://doi.org/10.58464/2155-5834.1211

Furlong, M.J., Greif, J.L., Bates, M.P., Whipple, A.D., Jimenez, T.C. & Morrison, R. (2005) Development of the California school climate and safety survey—short form. *Psychology in the Schools*, 42(2), 137–149. Available from: https://doi.org/10.1002/pits.20053

Furlong, M.J., Morrison, R. & Boles, S. (1991) *California school climate and safety survey*. Paper presented at the annual meeting of the California Association of School Psychologists, Los Angeles, CA.

Gustafson, J.-E. & Stahl, P.A. (2005) *STREAMS 3.0 user's guide*. Mölndal: MultivariateWare.

Healy, K. (2016) A theory of human motivation by Abraham H. Maslow (1942). *British Journal of Psychiatry*, 208(4), 313. Available from: https://doi.org/10.1192/bjp.bp.115.179622

Holley, L.C. & Steiner, S. (2005) Safe space: student perspectives on classroom environment. *Journal of Social Work Education*, 41(1), 49–64. Available from: https://doi.org/10.5175/JSWE.2005.200300343

Hong, J.S. & Espelage, D.L. (2012) A review of research on bullying and peer victimization in school: an ecological system analysis. *Aggression and Violent Behavior*, 17(4), 311–322. Available from: https://doi.org/10.1016/j.avb.2012.03.003

Hox, J.J. (2002) *Multilevel analysis: techniques and applications*. Mahwah: Erlbaum.

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

Huang, F.L. (2017) *Conducting multilevel confirmatory factor analysis using R*. Working paper. Available from: http://faculty.missouri.edu/huangf/data/mcfa/MCFAinRHUANG.pdf

Hurford, D.P., Londskog, R., Cole, A., Jackson, R., Thomasson, S. & Wade, A. (2010) The role of school climate in school violence: a validity study of a web-based school violence survey. *Journal of Educational Research & Policy Studies*, 10(1), 51–77.

James, L.R., Demaree, R.G. & Wolf, G. (1984) Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, 69(1), 85–98. Available from: https://doi.org/10.1037/0021-9010.69.1.85

Johnson, S.L. (2009) Improving the school environment to reduce school violence: a review of the literature. *Journal of School Health*, 79(10), 451–465. Available from: https://doi.org/10.1111/j.1746-1561.2009.00435.x

Jorgensen, T.D., Pornprasertmanit, S., Schoemann, A.M. & Rosseel, Y. (2018) *semTools: Useful tools for structural equation modeling*. R package version 0.5-1. Available from: https://CRAN.R-project.org/package=semTools

Kim, E.U., Dedrick, R.F., Cao, C. & Ferron, J.M. (2016) Multilevel factor analysis: reporting guidelines and a review of reporting practices. *Multivariate Behavioral Research*, 51(6), 881–898. Available from: https://doi.org/10.1080/00273171.2016.1228042

Kleinert, H.L., Browder, D.M. & Towles-Reeves, E.A. (2009) Models of cognition for students with significant cognitive disabilities: implications for assessment. *Review of Educational Research*, 79(1), 301–326. Available from: https://doi.org/10.3102/0034654308326160

Kline, R. (2015) *Principles and practice of structural equation modeling*, 4th edition. New York: Guilford Press.

Lacoe, J. (2016) Too scared to learn? The academic consequences of feeling unsafe in the classroom. *Urban Education*, 55(10), 1385–1418. Available from: https://doi.org/10.1177/0042085916674059

Li, C. (2015) Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48, 936–949. Available from: https://doi.org/10.3758/s13428-015-0619-7

Maras, P., Demetre, J., Moon, A. & Tolmie, A. (2012) *Psychology and antisocial behaviour in schools*. British Journal of Educational Psychology Monograph Series II: Psychological Aspects of Education—Current Trends: Number 9. British Psychological Society.

Marzano, R.J. & Gaddy, B.B. (2005) *A handbook for classroom management that works*. Cheltenham, VIC: ASCD.

Muthén, L.K. & Muthén, B.O. (1998–2010) *Mplus user's guide*, 6th edition. Los Angeles: Muthén & Muthén.

Nijs, M.M., Bun, C.J., Tempelaar, W.M., De Wit, N.J., Burger, H., Plevier, C.M. et al. (2014) Perceived school safety is strongly associated with adolescent mental health problems. *Community Mental Health Journal*, 50, 127–134. Available from: https://doi.org/10.1007/s10597-013-9599-1

Oberle, E., Domitrovich, C.D., Meyers, D.C. & Weissberg, R.P. (2016) Establishing systemic social and emotional learning approaches in schools: a framework for schoolwide implementation. *Cambridge Journal of Education*, 46(3), 277–297. Available from: https://doi.org/10.1080/0305764X.2015.1125450

Phillips, D.A. (2007) Punking and bullying: strategies in middle school, high school, and beyond. *Journal of Interpersonal Violence*, 22(2), 158–178. Available from: https://doi.org/10.1177/0886260506295341

R Core Team. (2017) *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available from: http://www.Rproject.org

Raudenbush, S. & Bryk, A. (2002) *Hierarchical linear models: applications and data analysis methods*, 2nd edition. Thousand Oaks, CA: Sage.

Rom, R.B. (1998) 'Safe spaces': reflections on an educational metaphor. *Journal of Curriculum Studies*, 30(4), 397–408. Available from: https://doi.org/10.1080/002202798183549

Rose, C.A., Monda-Amaya, L.E. & Espelage, D.L. (2010) Bullying perpetration and victimization in special education: a review of the literature. *Remedial and Special Education*, 32(2), 114–130. Available from: https://doi.org/10.1177/0741932510361247

Rosseel, Y. (2012) Lavaan: an R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. Available from: https://doi.org/10.18637/jss.v048.i02

Ryan, A.M., Pintrich, P.R. & Midgley, C. (2001) Avoiding seeking help in the classroom: who and why? *Educational Psychology Review*, 13(2), 93–114. Available from: https://doi.org/10.1023/A:1009013420053

Ryan, R.M. & Deci, E.L. (2017) *Self-determination theory: basic psychological needs in motivation, development, and wellness*. New York: Guilford Press.

Schroeder, S. (2016) Surveillance, violence, and the marginalization of students of color. *Penn GSE Perspectives on Urban Education*, 13(1), 56–59.

Steffgen, G., Recchia, S. & Viechtbauer, W. (2013) The link between school climate and violence in school: a meta analytic review. *Aggression and Violent Behavior*, 18(2), 300–309. Available from: https://doi.org/10.1016/j.avb.2012.12.001

Svetina, D., Rutkowski, L. & Rutkowski, D. (2020) Multiple-group invariance with categorical outcomes using updated guidelines: an illustration using Mplus and the lavaan/semTools packages. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(1), 111–130. Available from: https://doi.org/10.1080/10705511.2019.1602776

Taylor, S.E. (2006) Tend and befriend: biobehavioral bases of affiliation under stress. *Current Directions in Psychological Science*, 15(6), 273–277. Available from: https://doi.org/10.1111/j.1467-8721.2006.00451.x

Thapa, A., Cohen, J., Guffey, S. & Higgins-D'Alessandro, A. (2013) A review of school climate research. *Review of Educational Research*, 83(3), 357–385. Available from: https://doi.org/10.3102/0034654313483907

Thomas, D.E., Bierman, K.L. & Powers, C.J. (2011) The influence of classroom aggression and classroom climate on

aggressive–disruptive behavior. *Child Development*, 82(3), 751–757. Available from: https://doi.org/10.1111/j.1467-8624.2011.01586.x

Van de Schoot, R., Lugtig, P. & Hox, J. (2012) A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486–492. Available from: https://doi.org/10.1080/17405629.2012.686740

Van der Helm, G.H.P., Moonen, X., Stams, G.J.J.M., Giesen, N., Van der Heide, E.S. & Matthys, W. (2011) Measuring inappropriate responses of adolescents to problematic social situations in secure institutional and correctional youth care: A validation study of the TOPS-A. *Journal of Interpersonal Violence*, 28(8), 1579–1595. Available from: https://doi.org/10.1177/0886260512468322

Van Rest, M.M., Matthys, W., Van Nieuwenhuijzen, M., De Moor, M.H., Vriens, A. & Schuengel, C. (2019) Social information processing skills link executive functions to aggression in adolescents with mild to borderline intellectual disability. *Child Neuropsychology*, 25(5), 573–598. Available from: https://doi.org/10.1080/09297049.2018.1495186

Willner, P., Bailey, R., Parry, R. & Dymond, S. (2010) Evaluation of executive functioning in people with intellectual disabilities. *Journal of Intellectual Disability Research*, 54(4), 366–379. Available from: https://doi.org/10.1111/j.1365-2788.2010.01249.x

Wissink, I.B., Dekovic, M., Stams, G.J.J.M., Asscher, J.J., Rutten, E.A. & Zijlstra, B.J.H. (2014) Moral orientation and relationships in school and adolescent pro- and antisocial behaviors: a multilevel study. *Journal of School Nursing*, 30(3), 216–225. Available from: https://doi.org/10.1177/1059840513497402

Wu, H. & Estabrook, R. (2016) Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, 81(4), 1014–1045. Available from: https://doi.org/10.1007/s11336-016-9506-0

How to cite this article: Van der Helm, G.H.P., Klapwijk, G., Roest, J.J., Kuiper, C.H.Z., Scholte, R.H.J. & Stams, G.J.J.M. (2023) Measuring safety perceptions of students with behavioural problems in special education: A validation study of the safe at school questionnaire. *British Journal of Special Education*, 00, 1–11. Available from: https://doi.org/10.1111/1467-8578.12494