

**Assessing classroom emotional climate in STEM classrooms: Developing and validating
a questionnaire**

LERI-D-20-00079

Barry J. Fraser, Felicity I. McLure and Rekha B. Koul

Corresponding author: Barry J. Fraser (✉)

Barry J. Fraser, Felicity I. McLure and Rekha B. Koul

Curtin University

Perth WA 6845

Barry J. Fraser

Curtin University

Perth WA 6845

Email: B.Fraser@curtin.edu.au

This research was funded by an Australian Research Council grant number: DP180103259.

Assessing classroom emotional climate in STEM classrooms: Developing and validating a questionnaire

Submitted: 15 April 2020 – Accepted:

Abstract

In an attempt to engage more students in Science, Technology, Engineering and Mathematics (STEM) subjects, schools are encouraged by STEM educators and professionals to introduce students to STEM through projects which integrate skills from each of the STEM disciplines. Because little is known about the learning environment of STEM classrooms, we developed and validated a Classroom Emotional Climate (CEC) questionnaire. Initially, the questionnaire was pilot tested with six focus groups of students from three schools to obtain feedback to incorporate into a revision of the CEC. Next, the modified CEC questionnaire was administered to 698 students participating in STEM activities in 57 classes in 20 schools. Exploratory factor analysis (principal component analysis) led to reduction of the CEC to 41 items in seven dimensions: Consolidation, Collaboration, Control, Motivation, Care, Challenge and Clarity. The structure of the CEC was then further explored using confirmatory factor analysis. Internal consistency reliability, concurrent validity (ability to differentiate between classrooms), discriminant validity (scale intercorrelations) and predictive validity (associated with student attitudes) were satisfactory. Finally, Rasch analysis of data for each dimension revealed good model fit and unidimensionality of the items describing each latent variable.

Keywords: Attitudes, classroom emotional climate, emotions, learning environments, STEM education, validation

Introduction

“A historical look at the field of learning environment ... shows that a striking feature is the availability of a variety of economical, valid and widely-applicable questionnaires that have been developed and used for assessing students’ perceptions of classroom environment” (Fraser 1998, pp. 7-8). The evolution of this field and its importance both as an end in its own right and as a means for promoting educational outcomes are reviewed in Fraser (2014, 2019). An important gap in the field, however, is the lack of an economical, multidimensional and valid questionnaire for assessing students’ perceptions of classroom emotional climate.

The study reported in this article is distinctive within the learning environments field because, first, it provides such a questionnaire for measuring classroom emotional climate, second, it focuses on integrated STEM classes which currently are of considerable interest internationally and, third, our validation of the new questionnaire was extensive and encompassed exploratory and confirmatory factor analyses and Rasch analysis as well as several other techniques.

Although the central importance of science, technology, engineering and mathematics (STEM) is widely acknowledged internationally (Johnson et al. 2020), there are serious problems in terms of skills shortages in STEM professions and declining enrolments in secondary and tertiary STEM subjects (Office of Chief Scientist 2016), gender inequity in participation in STEM careers and enrolments (Koch, Polnick and Irby 2014) and declines in achievement levels in school science and mathematics in numerous countries according to PISA (OECD 2016) and TIMSS (Mullis et al. 2016). To alleviate these problems, currently

there is renewed international momentum in STEM school education, especially classes which integrate STEM domains in an attempt to increase student interest, awareness of the scope and importance of STEM for solving problems within society, achievement within individual STEM subjects and motivation to pursue careers within STEM (Honey, Pearson, and Schweingruber 2014). Other arguments for integrated STEM classes in schools are considered in the next section and include training students in transdisciplinary approaches to solve 21st century problems as preparation for many current and future careers (Nadelson and Seifert 2017).

With the introduction of integrated STEM education in schools, many questions arise about what factors within STEM classrooms support and encourage a positive emotional climate which prompts students to engage with STEM learning. Therefore, we developed and validated the Classroom Emotional Climate (CEC) questionnaire for obtaining feedback from students about the social-emotional climate within their STEM classrooms.

Background

Integrated STEM education

Nadelson and Seifert (2017) suggest that integrated STEM programs provide ill-structured problems whose multiple solutions require synthesis of skills from each of the STEM fields. These authors have categorised the STEM teaching domain into three categories ranging from segregated STEM subjects, for which the emphasis is on acquiring knowledge in each discipline, to fully integrated STEM programs, which involve problem-based or inquiry learning that entails the application of understanding and skills from each discipline and the use of higher-order thinking to arrive at solutions to those problems (Nadelson and Seifert

2017). Thibaut et al. (2018) identified five key principles in the literature that define integrated STEM classes: integration of STEM content; problem-centred learning; inquiry-based learning; design-based learning; and cooperative learning (which combines collaboration in small groups with teacher guidance).

Carefully-designed, authentic learning experiences within integrated STEM classrooms have the potential to enhance students' motivation, interest and experience of collaboration, as well as to provide a safe environment for putting forward and testing ideas, being challenged to think more deeply and synthesising ideas from all STEM domains (Nadelson and Seifert 2017). Studies suggest that participation in integrated STEM projects fosters student interest and affective, behavioural and cognitive engagement (Sinatra, Mukhopadhyay, and Allbright 2017). Additionally, STEM teachers' provision of guidelines and structure has been found to be associated with student motivation and engagement (Loof, Struyf, Pauw, and Petegem 2019). Also, classes in which students are undertaking STEM projects, such as the Creativity in Science and Technology (CREST) project developed in the United Kingdom, have had positive effects on students' motivation towards pursuing STEM subjects in secondary school and careers in STEM (Means, Wang, Young, Peters, and Lynch 2016; Moote 2019; Moote, Williams, and Sproule 2013).

Classroom Emotional Climate

Integrated STEM classrooms provide students with quite different learning experiences from those in segregated STEM subjects. Because types of integrated STEM classrooms are many and varied (Honey et al. 2014), questions arise about what factors within integrated STEM learning environments foster students' interest and motivation in pursuing STEM careers. Research into learning environments over several decades (Fraser 2012, 2014) has

convincingly established that the educational environment is a consistent determinant of students' cognitive and affective outcomes. This history of the field of learning environments is overviewed by Fraser (2012, 2019), who identifies the pioneering work of Walberg and Anderson (1968) and Moos and Trickett (1974) as the field's beginnings. Numerous economical, valid and widely-applicable questionnaires have been developed, validated and used for assessing students' perceptions of teacher–student interactions (Lee, Fraser and Fisher 2003) and the learning environments of science laboratory classrooms (Fraser, Giddings, and McRobbie 1995), constructivist-oriented classroom settings (Taylor, Fraser, and Fisher 1997) and outcomes-focused learning environments (Aldridge and Fraser 2008).

In addition to teacher–student interactions, emotional factors have become an area of focus more recently (Pekrun and Linnenbrink-Garcia 2014). The quality of the classroom emotional climate (Daniels and Shumow 2003; Pianta, La Paro, and Hamre 2008), or the social and emotional interactions between students and their peers and teacher, have been found to influence students' learning outcomes (Reyes, Brackett, Rivers, White, and Salovey 2012). The social and emotional atmosphere in which learning occurs also can influence students' motivation to engage in learning within that classroom environment (Urduan and Schoenfelder 2006), attitudes towards STEM and the desire to continue to study STEM subjects in later years (Khine 2015; Kind, Jones, and Barmby 2007).

Instruments measuring classroom emotional climate

According to Hamre and Pianta (2007), teachers in classrooms with positive classroom emotional climates: show care and concern for students; understand the needs of students; listen to students' points of view and take them into account; avoid sarcasm or harsh discipline; express interest and respect towards students and foster cooperation between

students; and are aware of students' emotional and academic needs. One instrument for determining the quality of classroom emotional climate, the Classroom Assessment Scoring System (CLASS, Pianta et al. 2008), is an observational tool for grades K–3 that focuses on the three domains of emotional support, classroom organisation and support for learning through instruction. Within the CLASS framework, classroom organisation is the teacher's ability to bring students' attention and behaviour to bear on academic activities, instructional support involves challenging and giving feedback to students so that they understand and investigate on a deeper level, and emotional support is the care, concern and respect that the teacher displays towards students and their points of view (Pianta and Hamre 2009).

The CLASS protocol involves 30-minute observations over several class sessions by trained observers. Use of the CLASS observation protocol in secondary schools has confirmed observations in primary classes that a positive classroom climate (in which the teacher is sensitive to students' needs and perspectives, uses varied teaching strategies, engages students' interests and challenges students by focusing on analysis and problem solving) can lead to higher levels of achievement (Allen et al. 2013).

While much insight has been gained into classroom emotional climate through classroom observations, they require extensive time and personnel resources. As the primary educational consumers who are present every day in the classrooms, students have very-valuable perspectives that require fewer resources to obtain. Multiple studies have shown that students' perceptions of their learning environment are reliable (e.g. Fauth, Decristan, Rieser, Klieme, and Buttner 2014; Wagner, Gollner, Helmke, Trautwein, and Ludtke 2013) and linked to students' motivation to learn (Spearman and Watt 2013).

Following on from the CLASS protocol, the Tripod student survey (Tripod 7 Cs) was developed to measure classroom emotional climate based on students' perspectives (Ferguson 2010). The Tripod's dimensions are: *care* (how much the teacher fosters a caring

environment where students feel they belong); *confer* (how much the teacher encourages students to express themselves and their points of view); *captivate* (how interesting and relevant the class is); *clarify* (recognising when students don't understand and providing alternative ways of teaching); *consolidate* (helping students to organise their knowledge and prepare for learning in the future); *challenge* (the teacher has high expectations for outcomes and encourages deeper thinking); and *control* (how well the teacher ensures that an environment is conducive to students focusing on tasks) (Ferguson 2012; Ferguson and Danielson 2014).

Past validations of 7 Cs instrument

Although the Tripod 7 Cs questionnaire has been widely used to evaluate instructional quality with almost a million students between 2001 and 2012 (Ferguson 2012), surprisingly little evidence has been published to support its validity. Theoretical models suggest either 3 or 7 dimensions as described by Pianta et al. (2008) and Ferguson (2010). Alternatively, the 7 dimensions of the Tripod 7 Cs questionnaire could describe the two more-fundamental dimensions of classroom experience of *academic press* and *social support for learning* (Ferguson and Danielson 2014). Whereas academic press involves factors which promote academic success, such as students being kept on task and encouraged to think deeply and persevere (the challenge and control dimensions of 7 Cs), social support encompasses social factors which promote trust, confidence and psychological safety within the classroom (the other 5 dimensions of the 7 Cs) (Ferguson and Danielson 2014).

A preliminary multi-level factor analysis of Tripod item responses from more than 25,000 students, however, failed to support the survey's theoretical 7-, 3- or 2-dimensional structure, but suggested an alternative bi-factor structure based on a general responsiveness

dimension and a classroom management dimension (Wallace, Kelcey, and Ruzek 2016). This was further confirmed by exploratory factor analysis for large samples from English and mathematics classes ($N > 7,000$), which supported a two-dimensional structure with control as one dimension and the other 6 Tripod factors making up a second dimensions of academic support (Kuhfeld 2017). Composite measures from the 7 Cs survey were shown by Rowley, Phillips, and Ferguson (2019) to be more stable than value-added or observational measures of classroom emotional climate across multiple grade levels.

Our questionnaire for assessing classroom emotional climate in STEM classrooms

The goal of our study was to develop and validate a widely-applicable and economical instrument to assess students' perceptions of classroom emotional climate in integrated STEM settings. Because little is known about either the impact of participation in integrated STEM projects on classroom emotional climate or the links between classroom emotional climate and students' attitudes/satisfaction in STEM classes, we developed the Classroom Emotional Climate (CEC) questionnaire. As a starting point, items from the CLASS observation protocol and the Tripod 7 Cs student perceptions survey were adapted for use in STEM classrooms to assess seven dimensions that we remade as care, control, clarity (clarify), challenge, motivation (captivate), consultation (confer) and consolidation (consolidate) (Ferguson 2012; Pianta et al. 2008). We added an eighth dimension, namely, students' experiences with collaboration in small groups, which is an important aspect of integrated STEM classes (Thibaut et al. 2018) to our CEC questionnaire. In order to allow investigation of the CEC's predictive validity, a STEM Attitudes questionnaire was developed based on selected items adapted from the Test of Science Related Attitudes (TOSRA, Fraser 1981).

Methods for questionnaire modification and validation

Below we describe our main three methods for refining and validating this questionnaire: face validation by experts; focus-group interviews with students; and a large-scale field test in schools followed by rigorous statistical analyses. Ethics approval for research on human subjects was obtained both from our university and state education department.

Phase 1: face validity

When the eight-dimension CEC questionnaire and the Attitudes questionnaire were examined by four experts in STEM education, learning environments research and questionnaire development for face validity, changes were made to item wording to improve clarity and readability. Additionally, all items were worded in a personal form rather than a class form as recommended by Fraser, Giddings and McRobbie (1995). For example, the Control item *Students in this class treat the teacher with respect* was modified to *I treat the teacher with respect* and the Care item *My teacher gives us time to explain our ideas* was changed to *My teacher gives me time to explain my ideas*.

Phase 2: focus groups

Four groups of grades 9 and 10 students ($N=24$) in integrated STEM classes in two Western Australian independent schools completed the initial form of the CEC questionnaire (8 dimensions and 64 items) and the Attitudes questionnaire (1 dimension and 8 items) before participating in focus-group discussions. Students were asked to give their opinions about the

applicability of questionnaire items to their underlying constructs (Care, Control, etc.) and to their experiences of integrated STEM classes. Students made suggestions about the deletion of unnecessary or similar items, the rewording of items that were ambiguous, and the addition of items to address aspects of each dimension that they believed were missing. Consensus was obtained within each focus group before modifications or changes were made. For instance, students unanimously felt that being shown respect by the teacher was very important for demonstrating that the teacher cared for them.

Based on recommendations of students from the four focus groups, changes were made to the original CEC questionnaire before a revised version was administered to two further groups of grades 9 and 10 integrated STEM students ($N=10$) from a different independent school. These students then discussed the questionnaire items in two focus groups and suggested some minor changes and additions. Consensus was again obtained within each focus group before modifications or changes were made to create an expanded questionnaire with 8 dimensions and 74 items and an Attitudes questionnaire with 1 dimension and 10 items. Sample items for each dimension are provided in Table 1.

Table 1 about here

Phase 3: large-sample validation

We contacted all government and independent schools in Western Australia which had classes that involved either integrated STEM or STEM projects as a component of a science subject to invite participation in completing the revised CEC (74 items) and Attitudes questionnaires (10 items) either online or in paper format. This led to a convenience sample of those schools and parents willing for students to complete the questionnaire. The sample

comprised students from 20 high schools and 58 STEM classes (either integrated or with STEM projects) ($N=698$: Male=294, Female=356, Unidentified=48). Eleven of the schools were non-government independent schools ($n=407$) and nine schools were government schools ($n=291$). There were 15 grade 7 classes, 9 grade 8 classes, 15 grade 9 classes, 7 grade 10 classes and 12 mixed grade-level classes. Class data were grouped through codes and all students completed the questionnaires anonymously. Human research ethics approval was obtained from both Curtin University and the Education Department of Western Australia and students were only included as participants with the written consent of teachers, parents and students.

Because earlier versions of the 7 Cs questionnaire previously had never been adequately validated, and also because of the numerous changes that we made to enhance its suitability for integrated STEM classrooms, we initially conducted exploratory factor analysis. Based on Stevens (2009), we chose principal component analysis (PCA) with direct oblimin rotation and Kaiser normalisation because of the expected correlation between dimensions.

Prior to conducting principal component analysis (PCA), the suitability of the data for factor analysis was checked using the Kaiser-Meyer-Olkin (KMO) test, with adequate sampling indicated by values between 0.8 and 1.0 (Cerny and Kaiser 1977). The criteria for retaining items were that they loaded 0.4 or higher on their own scale and less than 0.4 on all other scales. Dimensions with eigenvalues of 1.0 or greater were considered to contribute to the structure of the questionnaire (Stevens 2009). Because the initial CEC questionnaire consisted of a large number of items, we decided to reduce the questionnaire's length to a more-economical set of items for future use. In order to determine the suitability of the initial CEC questionnaire for item reduction, Bartlett's test for sphericity was carried out to compare the correlation matrix with the identity matrix, with a statistically-significant difference

indicating scope for item reduction (Bartlett 1950). Items with the lowest factor loadings were then removed from each dimension to obtain the final version of the CEC questionnaire.

The final measurement model in Figure 1 was evaluated by confirmatory factor analysis (CFA) using LISREL 10.20 (Jöreskog and Sörbom 2018) to determine goodness-of-fit indices for comparing the theoretical model with data from a polychoric correlation matrix (that was used because the questionnaire data were considered to be ordinal rather than interval). Because of the large number of observed variables/items (41), the number of responses collected was insufficient to generate an asymptotic covariance matrix, suggesting that some of the standard errors and χ^2 values could be unreliable (Jöreskog and Sörbom 2018).

Numerous indices of model fit were calculated. With a large number of data points, χ^2/df rather than χ^2 was used as recommended (Byrne 1998). Altogether, we used the following four model fit indices together with cut-off values recommended by Hair et al. (2020) and Alhija (2010): a small value of χ^2/df of <3.0; a root mean square error of approximation (RMSEA) of <0.08; a standardised root mean square residual (SRMR) of <0.06; and a comparative fit index (CFI) > 0.90.

Figure 1 about here

Once the latent variables assessed by the CEC questionnaire were confirmed using exploratory and confirmatory factor analysis, we checked internal consistency reliability using Cronbach's alpha coefficient (an example of convergent validity), concurrent validity (ability to differentiate between classrooms using ANOVA with class membership as the independent variable), discriminant validity (scale intercorrelations) and predictive validity

(associations with student attitudes using simple correlation and multiple regression analyses).

Finally, Rasch analysis of each CEC dimension was carried out to test its unidimensionality and reliability by transforming ordinal item scores into interval data and comparing both persons and items using a linear scale based on logits (Bond and Fox 2007; Wright 1999). Responses for items of each CEC dimension were analysed using Winsteps 4.4.7 software (Linacre, 2019) using the Rasch Rating Scale model for categorical data (Rasch 1960). We carried out Rasch person and item fit analyses and Rasch Principal Component Analysis (PCA) of item residuals and estimates of disattenuated person measures for each CEC dimension.

Infit and outfit mean square statistics measure the fit of items and persons to the latent variable being posited (Linacre 2002). Infit statistics are of most interest to researchers because they apply to most participants and are weighted towards persons who are near the endorsability level of the items, whereas outfit statistics are more affected by outliers (Bond and Fox 2007). Person and item misfit to the model can be caused by a small number of careless responses that can be examined further to determine whether they are unduly influencing the fit model fit (Wolfe and Smith 2007). In this study, a misfitting item was defined as one for which the mean square value was less than 0.5 or greater than 1.5 (Linacre 2019).

The reliability of each latent variable or dimension was estimated using person reliability (consistency of person responses) and person separation (ability of the dimension to separate persons into different response groups) (Wright and Masters 2002). A dimension with person separation >1.5 and person reliability >0.5 discriminates between two separate levels or groups of responders (e.g. high and low); person separation >2.0 and reliability >0.8 discriminates between three groups; and person separation >3.0 and reliability >0.9

discriminates between at least four groups (Boone, Staver, and Yale 2014; Linacre 2019). Person reliability is a more conservative measure than Cronbach's alpha coefficient (Linacre 1997). Item reliability (measuring coherence of items) and separation (ability of participants to distinguish between items) were also measured (Linacre 2019).

For each CEC dimension, a Rasch PCA of item residuals was conducted to determine the unidimensionality of the items within that construct. The observed raw variance explained by item measures of each dimension should be close to the expected values and preferably explain 50% or more of the total variance (Linacre 2007). The unexplained variance in the first PCA contrast should have an eigenvalue of less than 3.0 or explain less than 10% of the variance (Linacre 2007). In cases for which the unexplained variance is greater than 10%, it should be compared with the explained variance attributable to item difficulty and should not be greater than this value (Linacre 2019). A principal contrast with a greater eigenvalue and contrast than this suggests the presence of another construct in the latent variable.

Disattenuated person measures are questionnaire responses which have been transformed by the Rasch model to yield interval data. Correlation of disattenuated person measures for positively-loaded and negatively-loaded items from the Rasch PCA of item residuals can be used to further verify construct validity (Smith 2002). A high correlation (>0.70) suggests that all items are measuring the same dimension. Linacre (2019) suggests that a cut-off of $r=0.57$ indicates that person measures of the two item clusters share half of their variance. We examined the approximate disattenuated person measure correlations where unexplained variance in the first contrast was higher than 10% in order to further examine the unidimensionality of each dimension.

Results

Only the responses of the 660 students with complete answers to all CEC and Attitudes items were retained for data analyses. In reporting validity results below, four separate subsections are devoted to: exploratory PCA; CFA; Rasch analysis; and a set of other indicators (internal consistency reliability, concurrent validity, discriminant validity and predictive validity).

PCA (principal component analysis)

Prior to PCA of the 74-item CEC questionnaire and 10-item Attitude questionnaire, the factorability of each questionnaire was determined. The KMO (Kaiser-Meyer-Olkin) test for sample adequacy was 0.99 for the CEC and 0.96 for the Attitudes questionnaire, which indicates that sampling is more than adequate. Bartlett's test for sphericity for the CEC questionnaire was significant ($\chi^2(2701)=48023.50, p<0.001$) and indicated that reduction of items was appropriate.

Based on the proposed theoretical models of latent variables of the CEC questionnaire (Ferguson and Danielson 2014; Pianta et al. 2008) and validation studies of the Tripod 7 Cs (Kuhfeld 2017; Wallace et al. 2016), PCA was carried out for solutions with 2, 3, 4 and 8 dimensions. After direct oblimin rotation and Kaiser normalisation, examination of the resultant pattern matrix indicated poor structural fit for each of these solutions. For the eight-dimension solution, items from the Consultation and Consolidation dimensions had factor loadings of greater than 0.4 on the same dimension, suggesting that a seven-dimension structure was warranted, with Consultation and Consolidation items describing the same dimension. Although only five dimensions had eigenvalues of >1.0 , PCA with a five-dimension solution did not result in clearly-defined latent variables. For this reason, a seven-factor solution was chosen because the resultant factor loadings were mostly high and related to the theoretically-derived latent variables. In order to reduce the number of items describing

each latent variable, the pattern matrix was examined to identify any items with either low factor loadings on their own scale or high loadings on other dimensions. Multiple PCA checks throughout this process ensured that factor loadings remained high for each dimension. The final CEC questionnaire consists of 41 items, six for each dimension except Control, which has five items with factor loadings greater than 0.4. Factor loadings, communalities, eigenvalues and percentages of variance are presented in Table 2.

Table 2 about here

All CEC items met the criteria of having an absolute factor loading of >0.4 on the expected latent variable and <0.4 on other dimensions. Communalities of items were all greater than 0.3, indicating shared common variance with other items. The total proportion of variance explained by the seven latent variables was 74.69%. Eigenvalues for Challenge and Clarity of 0.84 and 0.71 indicate that they explained only a small amount of the total variance observed in the sample. The majority of the variance (52.547%) was explained by the combined Consultation and Consolidation dimensions, now named Consolidation, although items from both of the original item sets were included.

When the invariance of CEC questionnaire's structure for males ($n=278$) and females ($n=333$) was checked using PCA, the same seven-dimension structure was confirmed. All items satisfied the criteria for retention (a loading of at least 0.4 on their own scale and less than 0.4 on all other scales) for male students for all scales and for all items except two in the Clarity scale for females. The total proportion of variance explained by all 7 CEC factors was 75.57% for males and 75.65% for females.

Confirmatory factor analysis – measurement model

CFA was carried out on the reduced 41-item CEC questionnaire for our sample of 660 students to determine the fit of items to the proposed model with seven latent variables depicted in Figure 1. The first item for each latent variable was used as a marker indicator to scale the latent factor and the factor loading was set to 1.0. After correlating the errors of items describing each latent variable, we used the four model fit indices listed in Table 3, namely, χ^2/df , root mean square error of approximation (RMSEA), standardised root mean square residual (SRMR) and comparative fit index (CFI). Table 3 shows, for each of these four fit indices, the value obtained in our study, a cut-off value recommended in the literature and authors who recommended the cut-off value. Table 3 shows that values obtained for every fit index satisfied the cut-off criteria, therefore indicating acceptable fit between the theoretical CEC questionnaire model and the data from our sample.

Table 3 about here

In the measurement model for the CEC questionnaire in Figure 1, standardised estimates of factor loadings and standard errors for each item from CFA are presented. All factor loadings were statistically significant ($p < 0.05$) and showed a strong association between each item and its latent variable. Although a number of high standard errors could be a result of multicollinearity between items, it is more likely to be explained by the small sample size which did not allow production of an asymptotic covariance matrix, thus leading to unreliable standard errors (Jöreskog and Sörbom 2018). In order to overcome this problem, further future data collection is required. Overall, the CFA supported the 7-dimension theoretical factor structure proposed for the 41 CEC items.

Internal consistency reliability and concurrent, discriminant and predictive validities

The internal consistency reliability of each CEC scale was assessed using Cronbach's alpha coefficient. Table 4 shows that, for our sample of 660 integrated STEM students, alpha coefficients ranged from 0.91 to 0.95 for different CEC scales, which represents very high reliability according to Cohen, Manion and Morrison (2018).

Table 4 about here

A desirable characteristic of any classroom environment scale is that students within the same class have relatively similar perceptions while mean class perceptions vary from class to class. The concurrent validity of each CEC scale was investigated using an ANOVA with class membership as the independent in order to gauge the scale's ability to differentiate between classrooms. The ANOVA results in Table 4 indicate that all CEC scales were able to differentiate significantly ($p < 0.01$) between the perceptions of students in different classrooms and that the value of η^2 (which represents the proportion of variance attributable to class membership) ranged from 0.22 to 0.40 for different scales.

The discriminant validity (or independence) of the CEC scales was checked by examining the intercorrelations between scales, whose magnitudes ranged between 0.19 and 0.58. In Table 4, this information about scale intercorrelations is conveniently summarised for each scale by calculating the mean magnitude of the correlation of a scale with the other 6 scales. These values of the mean correlation in Table 4 range from 0.32 to 0.49, which are considerably smaller than scale reliability values, reflect satisfactory discriminant validity.

The predictive validity of the CEC was investigated in terms of associations between CEC scales and the 10-item Attitudes scale discussed in the Methods section. For the present sample, this Attitudes scale had a Cronbach alpha reliability coefficient of 0.98 and PCA confirmed the scale's unidimensionality with factor loadings ranging from 0.88 to 0.94, an eigenvalue of 8.41 and a proportion of variance accounted for of 84.06%. Both simple correlation and multiple regression analyses were used to explore, respectively, bivariate and multivariate associations between Attitudes and classroom emotional climate.

Table 4 reports correlations and standardised regression coefficients, which describe associations between Attitudes and a particular CEC scale when all the other CEC scales are mutually controlled. The association between Attitudes and a CEC scale was statistically significant for every scale for the simple correlation analysis and for the four scales of Collaboration, Motivation, Challenge and Clarity for the multiple regression analysis (Table 4). These associations between students' attitudes and their classroom environment perceptions replicate considerable prior research (Fraser 2019).

Rasch analysis of dimensions

Rasch analysis for the set of items comprising each latent CEC variable/scale revealed the results in Table 5 for item fit, person fit, reliability and unidimensionality. Because the length of the questionnaire deterred some students from completing all items, the sample size varied somewhat for different scales.

Table 5 about here

Infit and Outfit mean square values of between 0.50 and 1.50 indicated good fit for all dimensions with the exception of Motivation Item 6, which had an Infit value of 1.57. As suggested by Wolfe and Smith (2007), items were examined for misfitting individual item responses and, after removal of 16 such responses, Motivation Item 6 showed acceptable infit and outfit mean square values.

Person reliability was high (>0.80) for all dimensions except Control (0.78) and person separation values suggested that all latent variables except Control were able to discriminate between at least 3 levels of students. PCA of item residual data for all dimensions is presented in Table 5. The observed proportion of raw variance explained by measures was high ($>61\%$) in all cases and very similar to the expected proportion. The largest difference between observed and expected proportions of raw variance was for Control. Linacre's (2007) rule-of-thumb for unexplained variance in the first contrast having an eigenvalue of less than 3.00 was met, with only Control and Motivation having proportions of unexplained observed variance in the first contrast of slightly greater than 10%. This could suggest the presence of more than one dimension. However, examination of disattenuated person measures in both cases showed high correlations of 0.90 for Control and 0.85 for Motivation for positively- and negatively-loaded items from the Rasch PCA of item residuals, which is well above the recommended cut-off correlation of 0.57 (Linacre 2019). This indicates that the items are strongly correlated and can be considered as describing the same dimension.

Discussion and conclusion

This study filled an important gap within the field of learning environments by developing and validating a multidimensional and economical instrument for assessing students'

perceptions of classroom emotional climate. Our research also is distinctive in that it involved integrated STEM classrooms, which are currently of wide international interest. As well, our validation of this questionnaire was extensive and encompassed exploratory factor analysis, confirmatory factor analysis and Rasch analysis, as well as several other techniques (internal consistency reliability, concurrent validity, discriminant validity and predictive validity).

Modifications to the original version of the Classroom Emotional Climate (CEC) questionnaire and the evolution of a new version involved three phases. In the first phase, a group of experts in learning environments, questionnaire development and STEM education checked face validity and recommended modifications. The second phase involved interviews initially with 24 year 9 and 10 students in four focus groups and then, following changes, two further focus groups involving 10 students. In the third phase, after the CEC questionnaire was administered to a sample of grades 7–9 students in 59 STEM classes in 20 high schools, various statistical analyses of complete data from 660 students guided further refinements to the questionnaire.

Exploratory factor analysis (principal components analysis with direct oblimin rotation and Kaiser normalisation) yielded an optimal solution for 41 items in the 7 scales of Consolidation, Collaboration, Control, Motivation, Care, Challenge and Clarity. Each of these 41 items had a factor loading whose magnitude was at least 0.4 on its own scale and less than 0.4 on each of the other six scales. The eigenvalue for each of the seven scales was greater than or close to 1, and the total portion of variance accounted for was approximately 75%. Furthermore, the CEC questionnaire's structure was invariant to student gender, with the same seven-scale structure confirmed for males and females.

Confirmatory factor analysis of the 41-item CEC questionnaire revealed satisfactory fit between the data from 660 students and our measurement model with seven latent variables. The values obtained for four fit indices ($\chi^2/df = 2.92$, RMSEA = 0.07, SRMR=0.05,

CFI = 0.94) all satisfied cut-off criteria recommended by experts in the literature, therefore suggesting acceptable fit between a theoretical model for the questionnaire and our sample's data.

Rasch analysis for the set of items comprising each CEC variable/scale provided information about item fit, person fit, reliability and unidimensionality. Overall, nearly all obtained values were within recommended limits, indicating that items within each CEC scale are highly intercorrelated and describe a single dimension.

Evidence from four further types of analysis supported the validity of the CEC questionnaire. Internal consistency estimates (Cronbach alpha coefficients) ranged from 0.91 to 0.95 for different scales. Concurrent validity analyses revealed that each scale was capable of differentiating significantly between the perceptions of students in different classrooms. Discriminant validity indices (mean correlations of a scale with the other six scales) ranged from 0.22 to 0.40. The predictive validity of the questionnaire was supported by significant bivariate and multivariate associations between CEC scales and students' attitudes to STEM classes.

Like all educational research, our study had some possible limitations. Although the CEC questionnaire had been carefully developed and was relatively quick to administer, non-intimidating and easy to score, there are potential limitations to the questionnaire approach. Questionnaire items might have missed important aspects or failed to probe them in sufficient detail. Some students might have misinterpreted some items and therefore answered different questions from the ones intended. Other students might have distorted their answers because they felt uncomfortable about divulging certain information. Despite responses being anonymous, some students might have modified some answers to what they believed was expected.

Although our sample of 660 students was relatively large for a learning environment study, an even larger sample would have been associated with higher statistical power for statistical significance testing and more-accurate estimates of the underlying population. Relative to a broader and more representative sample, the results from our sample in Western Australia would not necessarily be generalisable to other groups. Our sample was too small to allow splitting into two separate subsamples, one for exploratory factor analysis and the other for confirmatory factor analysis. Also, because of the large number of individual items and latent variables/scales, sample limitations led to problems in obtaining the asymptotic covariance matrix and possibly to unreliable values for standard errors and χ^2 . Therefore, our confirmatory factor analysis results should be considered preliminary until replicated with other larger data sets.

The classroom emotional climate questionnaire validated in our study now can be used in the future by STEM teachers and researchers in several lines of learning environment research identified by Fraser (2019). These include the evaluation of educational programs (Lightburn and Fraser 2007) and the investigation of associations between the learning environment and student outcomes (Fraser and Fisher 1982), gender differences in perceived classroom emotional climate (Parker, Rennie and Fraser 1996), differences between students' and teachers' perceptions (Fraser 1984) and links between different environments (e.g. home and school environments, Fraser and Kahle 2007).

Another desirable direction would be to adapt the *actual* form of the CEC questionnaire (what the classroom is currently like) to create and validate a *preferred* form (what students would prefer the classroom to be like) (Fraser and Fisher 1983a). Having separate actual and preferred forms would permit: researchers to conduct person–environment fit studies of whether students achieve outcomes better in their preferred classroom environment (Fraser and Fisher 1983b); and practitioners to conduct action

research aimed at improving classrooms by matching actual to preferred environments (Fraser and Aldridge 2017; Aldridge, Fraser, Bell and Dorman 2012).

The limitations discussed above lead to suggestions for desirable future research. Including qualitative methods in future studies could lead to new insights and help to explain quantitative findings from the CEC questionnaire. As with any study, our sample had limitations that lead to a recommendation for larger future samples to provide greater statistical power and a more-diverse and more-representative sample to improve the generalisability of findings. Furthermore, although our study specifically involved STEM classrooms, we believe that our CEC questionnaire could be modified for use in a wide range of subject areas.

References

- Aldridge, J. M., & Fraser, B. J. (2008). *Outcomes-focused learning environments: Determinants and effects*. Rotterdam, Netherlands: Sense Publishers.
- Aldridge, J. M., Fraser, B. J., Bell, L., & Dorman, J. P. (2012). Using a new learning environment questionnaire for reflection in teacher action research. *Journal of Science Teacher Education, 23*, 259-290.
- Alhija, F. A. N. (2010). Factor analysis: An overview and some contemporary advances. In P. L. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (3rd ed., pp. 162-170). London: Elsevier.
- Allen, J., Gregory, A., Mikami, A., Lun, J., Hamre, B. K., & Pianta, R. C. (2013). Observations of effective teacher–student interactions in secondary school classrooms: Predicting student achievement with the classroom assessment scoring system–secondary. *School Psychology Review, 42*(1), 76-98.
- Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology, 3*, 77-85.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Erlbaum.
- Boone, W. J., Staver, J. R., & Yale, M. S. (Eds.). (2014). *Rasch analysis in the human sciences*. Dordrecht, Netherlands: Springer.
- Brown, T.A. (2014). *Confirmatory factor analysis for applied research* (2nd ed.). New York: Guilford.
- Byrne, B. (1998). *Structural equation modeling with LISREL, PRELIS and SIMPLIS: Basic concepts, application and programming*. Mahwah, NJ: Erlbaum.
- Byrne, B.M. (2002). *Structural equation modeling with EQS*. London: Taylor & Francis.

- Cerny, C. A., & Kaiser, H. F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate Behavioral Research*, 12(1), 43-47.
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). London: Routledge.
- Daniels, D. H., & Shumow, L. (2003). Child development and classroom teaching: A review of the literature and implications for educating teachers. *Journal of Applied Developmental Psychology*, 23, 495-526.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Buttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, 29, 1-9. doi:doi:10.1016/j.learninstruc.2013 .07.001
- Ferguson, R.F. (2010). *Student perceptions and the MET Project*. Bill & Melinda Gates Foundation. Retrieved from <https://k12education.gatesfoundation.org/resource/met-project-student-perceptions> on 4 April 2020
- Ferguson, R. F. (2012). Can student surveys measure teaching quality? *Phi Delta Kappan*, 94(3), 24-28.
- Ferguson, R. F., & Danielson, C. (2014). How framework for teaching and Tripod 7Cs evidence distinguish key components of effective teaching. In T. J. Kane, K. A. Kerr, & R. C. Pianta (Eds.), *Designing teacher evaluation systems* (pp. 98-143). San Francisco: Jossey-Bass.
- Fraser, B.J. (1981). *Test of Science Related Attitudes (TOSRA)*. Melbourne: Australian Council for Educational Research.
- Fraser, B.J. (1988). Classroom environment instrument: Development, validity and applications. *Learning Environments Research*, 1(1), 7-33.

- Fraser, B.J. (1984). Differences between preferred and actual classroom environment as perceived by primary students and teachers. *British Journal of Educational Psychology*, 54(3), 336-339.
- Fraser, B. J. (2012). Classroom learning environments: Retrospect, context and prospect. In B. J. Fraser, K. G. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (Vol. 2, pp. 1191-1239). New York: Springer.
- Fraser, B. J. (2014). Classroom learning environments: Historical and contemporary perspectives. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of research on science education* (Vol. II, pp. 104-117). New York: Routledge.
- Fraser, B.J. (2019). Milestones in the evolution of the learning environments field over the past three decades. In D.B. Zandvliet and B.J. Fraser (Eds.), *Thirty years of learning environments* (pp. 1-19). Leiden, The Netherlands: Brill | Sense.
- Fraser, B.J., & Aldridge, J.M. (2017). Improving classrooms through assessment of learning environments. In J.P. Bakken (Ed.), *Classrooms Volume 1: Assessment practices for teachers and student improvement strategies* (pp. 91-107). New York: Nova.
- Fraser, B. J., & Fisher, D. L. (1982). Predicting students' outcomes from their perceptions of classroom psychosocial environment. *American Educational Research Journal*, 19, 498-518.
- Fraser, B.J., & Fisher, D.L. (1983a). Development and validation of short forms of some instruments measuring student perceptions of actual and preferred classroom learning environment. *Science Education*, 67(1), 115-131.
- Fraser, B.J., & Fisher, D.L. (1983b). Student achievement as a function of person-environment fit: A regression surface analysis. *British Journal of Educational Psychology*, 53, 89-99.

- Fraser, B. J., Giddings, G. J., & McRobbie, C. J. (1995). Evolution and validation of a personal form of an instrument for assessing science laboratory classroom environments. *Journal of Research in Science Teaching*, 32, 399-422.
- Fraser, B.J., & Kahle, J.B. (2007). Classroom, home and peer environment influences on student outcomes in science and mathematics: An analysis of systemic reform data. *International Journal of Science Education*, 29, 189-1909.
- Hair, J.F., Jr., Black, W.C., Babin, B.J., & Anderson, R.E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Hamre, B. K., & Pianta, R. C. (2007). Learning opportunities in preschool and early elementary classrooms. In R. C. Pianta, M. J. Cox, & K. L. Snow (Eds.), *School readiness and the transition to kindergarten in the era of accountability* (pp. 49-83). Baltimore, MD: Brookes.
- Honey, M., Pearson, G., & Schweingruber, H. (Eds.). (2014). *STEM integration in K–12 education: Status, prospects, and an agenda for research*. Washington, DC: The National Academies Press.
- Johnson, C.C., Mohr-Schroeder, M.J., Moore, T.J., & English, L. (Eds.). (2020). *Handbook of research on STEM education*. London: Routledge.
- Jöreskog, K. G., & Sörbom, D. (2018). *LISREL10: User's reference guide*. Chicago: Scientific Software Inc.
- Khine, M. S. (Ed.) (2015). *Attitude measurements in science education: Classic and contemporary approaches*. Charlotte, NC: Information Age Publishing.
- Kind, P., Jones, K., & Barmby, P. (2007). Developing attitudes towards science measures. *International Journal of Science Education*, 29(7), 871-893.
- doi:10.1080/09500690600909091

- Koch, A., Polnick, B., & Irby, B. (Eds.). (2014). *Girls and women in STEM: A never ending story*. Charlotte, NC: Information Age Publishing.
- Kuhfeld, M. (2017). When students grade their teachers: A validity analysis of the Tripod student survey. *Educational Assessment*, 22(4), 1-22.
doi:10.1080/10627197.2017.1381555
- Lee, S.S.U., Fraser, B.J., & Fisher, D.L. (2003). Teacher-student interactions in Korean high school science classrooms. *International Journal of Science and Mathematics Education*, 1 (1), 67-85.
- Lightburn, M.E., & Fraser, B.J. (2007). Classroom environment and student outcomes among students using anthropometry activities in high school science. *Research in Science and Technological Education*, 25, 153-166.
- Linacre, J.M. (1997). KR-20 or Rasch reliability: Which tells the "truth"? *Rasch Measurement Transactions*, 11(3), 580-581.
- Linacre, J. M. (2002). What do infit, outfit, mean-square, and standardized mean? *Rasch Measurement Transactions*, 16.
- Linacre, J. M. (2007). *A user's guide to WINSTEPS: Rasch-model computer program*. Chicago: MESA.
- Linacre, J. M. (2019). *Winsteps (Version 4.4.7)*. Chicago: Winsteps.com.
- Loof, H., Struyf, A., Pauw, J., & Petegem, P. (2019). Teachers' motivating style and students' motivation and engagement in STEM: The relationship between three key educational concepts. *Research in Science Education*, 1-19. doi:10.1007/s11165-019-9830-3
- Means, B., Wang, H., Young, V., Peters, V. L., & Lynch, S. J. (2016). STEM-focused high schools as a strategy for enhancing readiness for postsecondary STEM programs. *Journal of Research in Science Teaching*, 53(5), 709-736.

- Moos, R.H., & Trickett, E.J. (1974). *Classroom Environment Scale manual*. Palo Alto, CA: Consulting Psychologists Press.
- Moote, J. K. (2019). Investigating the impact of classroom climate on UK school students taking part in a science inquiry-based learning programme—CREST. *Research Papers in Education*. doi:10.1080/02671522.2019.1568533
- Moote, J. K., Williams, J. M., & Sproule, J. (2013). When students take control: Investigating the impact of the CREST inquiry-based learning program on self-regulated processes and related motivations in young science students. *Journal of Cognitive Education and Psychology, 12*(2), 178-196.
- Mullis, I.V.S., Martin, M.O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 international results in mathematics*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website:
<http://timssandpirls.bc.edu/timss2015/internationalresults/>
- Nadelson, L. S., & Seifert, A. L. (2017). Integrated STEM defined: Contexts, challenges, and the future. *The Journal of Educational Research, 110*(3), 221-223.
doi:10.1080/00220671.2017.1289775
- OECD (Organisation for Economic Co-operation and Development). (2016). *PISA 2015 results (Volume I): Excellence and equity in education*. Paris: OECD Publishing.
- Office of Chief Scientist. (2016). *Australia's STEM workforce*. Canberra: Australian Government.
- Parker, L.H., Rennie, L.J., & Fraser, B.J. (Eds.). (1996). *Gender, science and mathematics: Shortening the shadow*. London: Falmer Press.
- Pekrun, R., & Linnenbrink-Garcia, L. (Eds.). (2014). *International handbook of emotions in education*. New York: Routledge.

- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109-119. doi:doi:10.3102/0013189X09332374
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System™: Manual K–3*. Baltimore, MD: Brookes.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago.
- Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology*, 104(3), 700-712. doi:10.1037/a0027268
- Rowley, J. F. S., Phillips, S. F., & Ferguson, R. F. (2019). The stability of student ratings of teacher instructional practice: Examining the one-year stability of the 7Cs composite. *School Effectiveness and School Improvement*, 30(4), 549-562
doi:10.1080/09243453.2019.1620293
- Sinatra, G. M., Mukhopadhyay, A., & Allbright, T. N. (2017). Speedometry: A vehicle for promoting interest and engagement through integrated STEM instruction. *The Journal of Educational Research*, 110(3), 308-316. doi:10.1080/00220671.2016.1273178
- Smith, E. V. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, 3, 205-231.
- Spearman, J., & Watt, H. G. M. (2013). Perception shapes experience: The influence of actual and perceived classroom environment dimensions on girls' motivations for science. *Learning Environments Research*, 16(2), 217-238. doi:doi:10.1007/s10984-013-9129-7

- Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). New York: Routledge.
- Taylor, P. C., Fraser, B. J., & Fisher, D. L. (1997). Monitoring constructivist classroom learning environments. *International Journal of Educational Research*, 27, 293-302.
- Thibaut, L., Ceuppens, S., De Loof, H., De Meester, J., Goovaerts, L., Struyf, A., . . . Depaepe, F. (2018). Integrated STEM education: A systematic review of instructional practices in secondary education. *European Journal of STEM Education*, 3(1).
doi:10.20897/ejsteme/85525
- Urduan, T., & Schoenfelder, E. (2006). Classroom effects on student motivation: Goal structures, social relationships, and competence beliefs. *Journal of School Psychology*, 44, 331-349.
- Wagner, D., Gollner, R., Helmke, A., Trautwein, U., & Ludtke, O. (2013). Construct validity of student perceptions of instructional quality is high, but not perfect: Dimensionality and generalizability of domain-independent assessments. *Learning and Instruction*, 28, 1-11. doi:doi:10.1016/j.learninstruc.2013.03.003
- Walberg, H. J., & Anderson, G. J. (1968). Classroom climate and individual learning. *Journal of Educational Psychology*, 59, 414-419.
- Wallace, T. L. B., Kelcey, B., & Ruzek, E. (2016). What can student perception surveys tell us about teaching? Empirically testing the underlying structure of the tripod student perception survey. *American Educational Research Journal*, 53(6), 1834-1868.
doi:10.3102/0002831216671864
- Wolfe, E. W., & Smith, E. V. (2007). Instrument development tools and activities for measure validation using Rasch models: Part II – Validation activities. *Journal of Applied Measurement*, 8, 204-234.

- Wright, B. D. (1999). Fundamental measurement for psychology. In E. S. E. & S. L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 65-104). Mahwah, NJ: Erlbaum.
- Wright, B. D., & Masters, G. N. (2002). Number of person or item strata. *Rasch Measurement Transactions*, 16.

List of figures

Fig 1. Measurement model for CEC questionnaire

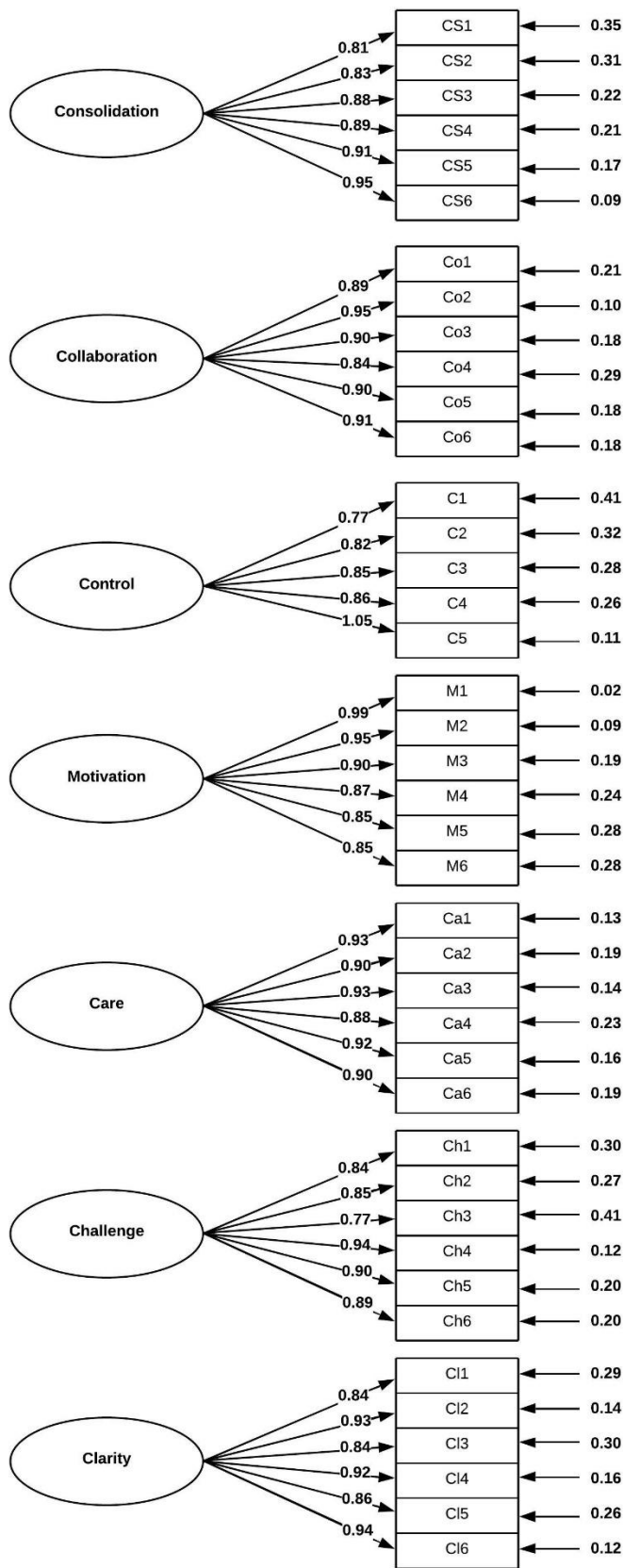


Fig. 1

Table 1 Scale description and sample items from initial CEC and Attitudes questionnaires

Scale	Scale description Extent to which students perceive that...	Samples items
Consultation	...their teacher encourages them to share their thoughts or opinions	My teacher asks me questions whether I put up my hand or not. I speak up and share my ideas about the class work.
Consolidation	...their teacher consolidates understanding and indicates how to improve	My teacher takes time to summarise what I learn each day. My teacher speaks to me in class about how to improve. My teacher points me in the right direction to get further help.
Collaboration	...they collaborate effectively with peers to complete tasks	When I work in a group doing a STEM project, we work as a team. I help other group members who are having trouble doing STEM projects. I learn from other students in my STEM group.
Control	...they behave well in class and remain focused on tasks	My teacher makes sure that I stay busy and don't waste time. I behave well when the teacher is explaining things to the class. I treat the teacher with respect.
Motivation	...their STEM project and the classroom environment motivates them to be engaged in learning and seeking answers to problems	The questions in this class make me want to find out the answers. How the classroom looks makes me motivated. My project makes me want to learn.
Care	...their teacher cares about their learning needs and treats them with respect	My teacher makes me feel that s/he really cares about me. My teacher gives me time to explain my ideas. My teacher encourages me to do my best.
Challenge	...their teacher encourages them to think deeply and persist in the face of difficulties and provides challenging STEM projects	My teacher asks questions that make me think hard. My teacher encourages me to keep going when the work is hard. I'm encouraged to correct my mistakes.
Clarity	...their teacher gives clear directions and explanations	My teacher uses a variety of teaching methods to make things clear to me. My teacher knows when I understand and when I do not. My teacher breaks up the work into easy steps.
Attitudes	Students' interest, enjoyment and confidence in doing STEM projects and their desire to continue studying STEM subjects	I feel confident about doing STEM projects in class. These lessons make me interested in studying a STEM subject in the future. I enjoy lessons that are part of STEM projects.

Table 2 Exploratory factor analysis (PCA) results for CEC questionnaire (41 items)

Item	Factor Loadings							Communalities
	Consolidation	Collaboration	Control	Motivation	Care	Challenge	Clarity	
1	0.67							0.63
2	0.72							0.73
3	0.62							0.72
4	0.70							0.75
5	0.67							0.78
6	0.64							0.81
7		0.85						0.77
8		0.85						0.80
9		0.86						0.78
10		0.78						0.66
11		0.87						0.78
12		0.82						0.74
13			0.87					0.79
14			0.68					0.65
15			0.82					0.77
16			0.76					0.70
17			0.46					0.67
18				-0.44				0.78
19				-0.53				0.77
20				-0.81				0.87
21				-0.85				0.86
22				-0.87				0.85
23				-0.63				0.70
24					-0.71			0.80
25					-0.84			0.83
26					-0.61			0.76
27					-0.71			0.72
28					-0.57			0.73
29					-0.78			0.78
30						-0.57		0.69
31						-0.52		0.67
32						-0.74		0.73
33						-0.54		0.77
34						-0.48		0.69
35						-0.47		0.68
36							-0.73	0.75
37							-0.61	0.80
38							-0.62	0.71
39							-0.49	0.74
40							-0.53	0.70
41							-0.48	0.78
% Variance	52.55	7.90	4.37	3.37	2.70	2.05	1.74	
Eigenvalue	21.54	3.24	1.79	1.39	1.11	0.84	0.71	

N=660 students in 20 high schools

Only factor loading above |0.4| are shown

Table 3 Model fit indices, obtained values and cut-off guidelines for CFA of CEC questionnaire

Model fit index	Obtained value	Cut-off guideline	References
χ^2/df	2.92	<3	Hair et al., 2010; Alhija 2010
RMSEA	0.07	<0.08	Hair et al. (2010); Alhija 2010
SRMR	0.05	<0.08	Hair et al. (2010); Alhija (2010)
CFI	0.94	>0.90	Byrne (2002); Brown (2014)

Table 4. Internal consistency reliability (alpha coefficient), concurrent validity (ANOVA results for ability to differentiate between classrooms), discriminant validity (mean correlation with other scales) and predictive validity (simple correlation and multiple regression results for associations with attitudes)

CEC scale	Alpha reliability	Concurrent validity	Discriminant validity	Predictive validity (Association with attitudes)	
		ANOVA (η^2)	Mean correlation	<i>r</i>	β
Consolidation	0.94	0.39**	0.46	0.61**	0.06
Collaboration	0.94	0.29**	0.41	0.54**	0.22**
Control	0.91	0.22**	0.32	0.46**	-0.02
Motivation	0.95	0.40**	0.48	0.76**	0.64**
Care	0.95	0.39**	0.48	0.56**	-0.09
Challenge	0.92	0.36**	0.46	0.60**	-0.11*
Clarity	0.93	0.38**	0.49	0.65**	0.15**
Multiple correlation				0.78**	

* $p < 0.05$, ** $p < 0.01$

Table 5 Rasch analysis of CEC scales

Dimension (<i>N</i>)	No of items	Person reliability	Person separation	PCA analysis of item residuals				
				Raw variance explained by measures			Unexplained variance in 1 st contrast	
				Eigenvalue	Observed %	Expected %	Eigenvalue	Observed %
Consolidation (673)	6	0.84	2.29	10.70	64.1	63.9	1.47	7.4
Collaboration (662)	6	0.83	2.22	9.93	62.3	62.3	1.55	10.7
Control (694)	5	0.78	1.86	9.19	64.8	65.7	1.52	7.9
Motivation (677)	6	0.87	2.60	12.87	68.2	68.2	2.10	9.7
Care (698)	6	0.84	2.32	13.76	69.6	70.4	1.76	11.1
Challenge (681)	6	0.83	2.20	9.69	61.8	61.7	1.52	8.8
Clarity (687)	6	0.86	2.51	12.75	68.0	67.7	1.47	9.7

