

‘Are We Included?’ A validation of the Teachers’ Inclusion Climate Scale

6th October 2022

**Elpida Achtaridou, Jason Lamprianou, Sonia
Blandford, Suanne Gibson, Jen Shute, George Munn
and Umesh Sharma**

**Marjon Evaluation & Research Impact Centre (MERIC), Plymouth
Marjon University
Institute of Education, Plymouth Marjon University
Plymouth Institute of Education, University of Plymouth**

Acknowledgements

We would like to thank Plymouth City Council for their continuous support throughout the lifecycle of this study. We would also like to thank all the Plymouth schools who took part in the study and particularly their remarkable teacher who enthusiastically filled in the inclusion questionnaire and provided invaluable insights during the interview process.

Contents

Section 1 - Introduction.....	3
Section 2 – Data and Methods.....	5
Section 3 – Results.....	11
Section 4 – Summary of Recommendations.....	41
Bibliography.....	42
Appendices.....	44

Section 1 - Introduction

In March 2022 Plymouth City Council commissioned Plymouth Marjon University and the University of Plymouth to conduct research on three assessment instruments measuring inclusion from the perspectives of students, parents and teachers.

The research projects were part of *Plymouth Place-based School Improvement* project. At the time the project focused on three strategic priorities: school leadership, curriculum and inclusion. To deliver on its priorities a number of key stakeholders were brought together, including the Education Endowment Fund (EEF), Plymouth Local Authority, the Regional Schools' Commissioner, Headteachers, MAT CEOs and their leadership teams who chosen to take a collaborative and sustainable approach to ensure rapid school improvement.

The research aimed to support Plymouth's strategic priority of inclusion and took place between May and August 2022. The concept of 'inclusion education' has been debated in Organisation for Economic Co-operation and Development (OECD) countries since an early UNESCO (1994) report. Since then, the debate has been enriched with a multitude of ideas, mainly championing the initiatives of comprehensive learning environments where all children would be included in the learning but also social life. Highly cited research, however, such as Avramidis and Norwich (2002) has maintained that any inclusive policy - no matter how well-designed or funded it is, depends heavily on the attitudes of teachers to be successful. Various factors have also been identified as affecting teachers' attitudes towards inclusion education, such as teachers' experience, school ethos etc.

Teachers, however, are not the single important dimension which affects the success of failure of any inclusion policy or implementation. In a very meticulous literature review, Qvortrup and Qvortrup (2018) have proposed additional important dimensions of consideration, such as the types of 'social communities' in and out of school, which may include the role of class context, the relationship of learners with other school agents such as teachers, other children, staff etc. Comprehensive considerations which take into account both teachers, learners and parents seem to be necessary in order to have a holistic view of inclusive education in any given learning environment.

Investigating the perceptions and role of learners has been intensified in the last two decades, mainly with quantitative surveys, but also with other research paradigms. For example, Schwab et al., (2018) conducted a relatively large-scale survey of students' (aged 10-17 years) perception of the climate in their classrooms, using the Inclusion Climate Scale (ICS). The findings revealed that there are two major dimensions of students' attitude: teacher support and emotional experience. This comes to no surprise and teachers' role is expected to be central in the every-day life of a young learner; certainly, the emotional experienced of a young learner would be expected to depend heavily on the degree of support an individual receives in the class. Sointu et al., (2017), for example, Sointu et al., (2017) provided evidence that "Positive student-teacher relationships are related to students' academic achievement and behavioural and emotional adjustment" (p.457).

Early on, however, it was found that parents - in addition to teachers - could be an important agent to facilitate successful inclusion. De Boer et. al., (2010), in a very informative literature review, suggested that a positive parental attitude towards inclusion is very important for children to enjoy a successfully inclusive life in the learning environment. Large scale research focusing on parents in the last years has

also re-iterated their important role for a successful implementation of inclusive education (see, for example, Paseka and Schwab, 2020).

As it has been discussed in the paragraphs above, we may consider teachers, learners and parents to be the vertices of an isosceles triangle. As a result, *the 'Are We Included?' A validation of the Teachers'* project has been developed, employing both qualitative and quantitative methods to investigate the role of all three important agents of inclusion education: parents, students and teachers.

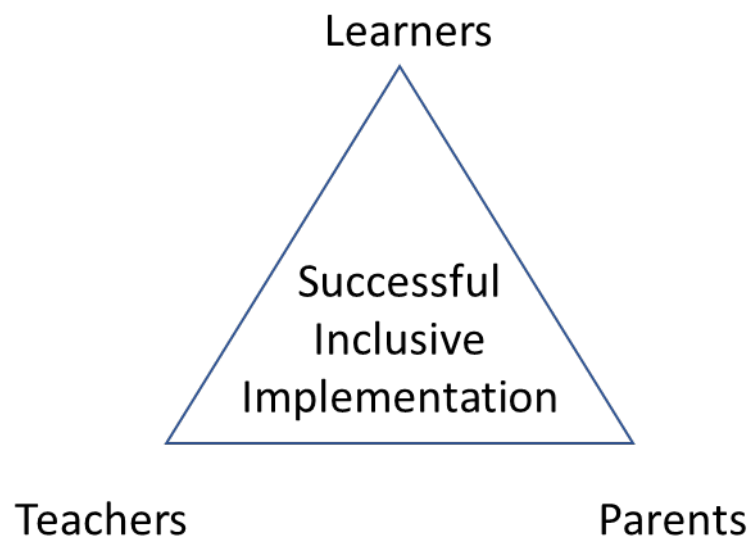


Figure 1. Three important agents of inclusion education: parents, students and teachers.

For each of the three distinct populations - parents, teachers and students - , the project collects data using different scales, based on past research. For example, to investigate teacher attitudes, the research uses scales which have been developed and tested relatively recently (for example, see Sharma, Loreman and Forlin, 2012; Sharma and Jacobs, 2016; Sharma et al., 2021).

This report focuses on the findings of the research on the teacher questionnaire. To survey teachers, the project used the same instrument which was originally developed and used by Schwab et al., 2022.

Section 2 – Data and Methods

Aims and Objectives

This study aimed to validate the individual scales of the teacher questionnaire measuring inclusion in the context of secondary schools in Plymouth, England

The study is important as it assesses the validity and reliability of the scales in the context of Plymouth where it is to be used, with possibly the prospect of piloting it in different schools in England at a later stage. Moreover, if our findings corroborate the findings of past research, the research community would be encouraged to use of the questionnaire in different school settings. Similar findings from different countries would suggest that the questionnaire produces valid and reliable results across educational systems and cultures.

Data and Methods

The instrument

A. Attitudes to Inclusion Scale (AIS)

The Attitudes to Inclusion Scale (AIS) measures teachers' attitudes to the inclusion of students with diversities in schools (regular schools, not special schools). This set of 8 items uses a 7-step scale: 1. Strongly disagree, 2. Moderately disagree, 3. Slightly disagree, 4. Undecided, 5. Slightly agree, 6. Moderately agree, 7. Strongly agree.

The Attitudes to Inclusion Scale (AIS) includes the following items:

1. I believe that all students regardless of their ability should be taught in regular classrooms.
2. I believe that inclusion is beneficial to all students socially.
3. I believe that inclusion benefits all students academically.
4. I believe that all student can learn in inclusive classrooms if their teachers are willing to adapt the curriculum.
5. I am pleased that I have the opportunity to teach students with lower academic ability alongside other students in my class.
6. I am excited to teach students with a range of abilities in my class.
7. I believe that including students with a range of abilities will make me a better teacher.
8. I am happy to have students who need assistance with their daily activities included in my classrooms.

The original scale used in past research (see the Literature Review section) had ten items:

1. "I believe that all students regardless of their ability should be taught in regular classrooms."
2. "I believe that inclusion is beneficial to all students socially."
3. "I believe that inclusion benefits all students academically."
4. "I believe that all student can learn in inclusive classrooms if their teachers are willing to adapt the curriculum."

5. "I believe that placement of students with severe disabilities in special schools is the best option for education of such students."
6. "I believe that students with social emotional behaviors should be taught in special schools."
7. "I am pleased that I have the opportunity to teach students with lower academic ability alongside other students in my class."
8. "I am excited to teach students with a range of abilities in my class."
9. "I am pleased that including students with a range of abilities will make me a better teacher."
10. "I am happy to have students who need assistance with their daily activities included in my classrooms."

Important note: Our version of the questionnaire does NOT include questions 5 and 6 of the original scale, which makes difficult the comparability of our results with those of Sharma & Jacobs (2016). Also, our version of the questionnaire includes a re-worded version of the item 9 of the original scale: the verb 'pleased' in "I am pleased that including students with a range of abilities will make me a better teacher." was changed to 'I believe'. It is not clear what problems this could create, as the item might be expected to load on the 'belief' rather than the 'feelings' factor: According to Sharma & Jacobs (2016), we expect the first items to load on a 'Beliefs' factor and the last four items to load on a 'Feelings' factor.

Interestingly, in Sharma, Sokal, Wang & Loreman (2021), the AIS scale was used as a unidimensional one (they used the sum/raw score), although in Sharma & Jacobs (2016) they showed that the scale consists of two sub-scales.

"An 18-item scale was developed on a sample of 607 pre-service teachers selected from four countries (Canada, Australia, Hong Kong and India). Factor analysis of responses from the sample revealed three factors: efficacy in using inclusive instruction, efficacy in collaboration and efficacy in dealing with disruptive behaviours. The alpha coefficient for the total scale was 0.89. Alpha coefficients for the three factors ranged from 0.85 to 0.93." Sharma, Loreman & Forlin (2012)

B. Intention to Teach in Inclusive Classrooms (ITIC)

This scale relates to the intention of teachers to teach students who need additional support. This set of items uses a 7-step scale: 1. Extremely unlikely, 2. Very unlikely, 3. Somewhat unlikely, 4. Not sure, 5. Somewhat likely, 6. Very likely, 7. Extremely likely:

The Intention to Teach in Inclusive Classrooms (ITIC) included the following items:

1. Change the curriculum to meet the learning needs of a student with learning difficulty enrolled in your class.
2. Consult with the parents of a student who is struggling in your class.
3. Consult with your colleagues to identify possible ways you can assist a struggling student in your class.
4. Undertake a professional development program so you can teach students with diverse learning needs well.
5. Consult with a student who is displaying challenging behaviours to find out better ways to work with him/her.
6. Include students with severe disabilities in a range of social activities in your class.

7. Change the assessment tasks to suit the learning profile of a student who is struggling (e.g., providing longer time to complete the task or modifying test questions).

For the ITIC scale, we used the same questions as those used by Sharma & Jacobs (2016). Judging by the findings of past research, we expect two factors, one for the 'Intentions for curriculum change' (items 1, 6, 7) and one for the 'Intentions to consult' (items 2, 3, 4, 5).

C. Teaching Efficacy to implement Inclusive Practices- TEIP

This scale is designed to uncover the opinion of teachers regarding the factors influencing the success of classroom activities in creating an inclusive classroom environment. This set of items uses a 6-step scale: 1. Strongly disagree, 2. Disagree, 3. Disagree somewhat, 4. Agree somewhat, 5. Agree, 6. Very likely, 7. Strongly agree

There are eighteen items in the scale:

1. I can use a variety of assessment strategies (for example, portfolio assessment, modified tests, performance-based assessment, etc.).
2. I am able to provide an alternate explanation or example when students are confused.
3. I am confident in designing learning tasks so that the individual needs of students with disabilities are accommodated.
4. I can accurately gauge what progress students have made in a taught area
5. I can provide appropriate challenges for very capable students.
6. I am confident in my ability to get students to work together in pairs or in small groups.
7. I am confident in my ability to prevent disruptive behaviour in the classroom before it occurs.
8. I can control disruptive behaviour in the classroom.
9. I am able to calm a student who is disruptive or noisy.
10. I am able to get children to follow classroom rules.
11. I am confident when dealing with students who are physically aggressive.
12. I can make my expectations clear about student behaviour.
13. I can assist families in helping their children do well in school.
14. If a student is failing I can adapt my teaching to help them succeed
15. I am able to work jointly with other professionals and staff (e.g. aides, other teachers) to teach students with disabilities in the classroom.
16. I am confident in my ability to get parents involved in school activities of their children with disabilities.
17. I can collaborate with other professionals (e.g itinerant teachers or speech pathologists) in designing educational plans for students with disabilities.
18. I am confident in informing others who know little about laws and policies relating to the inclusion of students with disabilities.

For the TEIP scale, we used the same questions as those used by Sharma & Jacobs (2016). Past research suggests that we could expect three factors: Efficacy in inclusive instruction (items 1-6), Efficacy in managing behavior (items 7-12), and Efficacy in collaboration (items 13-18).

D. Teacher Efficacy for Inclusive Practice – Collective (TEIP-C) Scale

This scale is very similar to the TEIP scale discussed above. It is designed to uncover the opinion of teachers regarding the factors influencing the success of classroom activities in creating an inclusive classroom environment. However, the

responders are asked to consider the overall (i.e., 'general' or 'average') capacities of the group of teachers who teach at the school (not the capacity of the responder).

It is not clear why the authors of the TEIP-C scale chose to change the order of the items and did not follow exactly the same format as the TEIP scale. This may probably lead to different findings regarding the psychometric characteristics of the scale.

This set of items uses a 6-step scale: 1. Strongly disagree, 2. Disagree, 3. Disagree somewhat, 4. Agree somewhat, 5. Agree, 6. Very likely, 7. Strongly agree

There are eighteen items in the scale:

1. Teachers in my school can make their expectations clear about desired student behaviour.
2. Teachers in my school are able to calm a student who is disruptive or noisy.
3. Teachers in my school can make parents feel comfortable coming to school.
4. Teachers in my school can assist families in helping their children do well in school.
5. Teachers in my school can accurately gauge what progress students have made in a taught area.
6. Teachers in my school can provide appropriate challenges for very capable students.
7. Teachers in my school can prevent disruptive behaviour in the classroom before it occurs.
8. Teachers in my school can control disruptive behaviour in the classroom.
9. Teachers in my school can get parents of children who are frequently excluded involved in school activities.
10. Teachers in my school design learning tasks so that the individual learning needs of all students are accommodated.
11. Teachers in my school get children to follow classroom rules.
12. Teachers in my school can collaborate with other professionals (e.g., itinerant teachers or speech pathologists) in designing educational plans for students with disabilities.
13. Teachers in my school are able to work jointly with other professionals and staff (e.g. aides, other teachers) to ensure that all students are included in the classroom.
14. Teachers in my school can get students to work together cooperatively in pairs or in small groups.
15. Teachers in my school can use a variety of assessment strategies in order to determine if all children in a class are learning (for example, portfolio assessment, modified tests, performance-based assessment, etc.).
16. Teachers in my school are confident in informing others who know little about laws and policies relating to the inclusion of students with disabilities.
17. Teachers in my school are confident in managing a situation if a student becomes physically aggressive in the classroom.
18. Teachers in my school are able to provide an alternate explanation or example when students are confused.

For the analysis, we used EFA to reveal the underlying factors in the data. Due to the ordinal nature of the scale (e.g. "Not at all true" to "Completely true" or other similar format), we used polychoric correlations to compute the correlation matrix for the EFA.

We used standard psychometric techniques (e.g. item-total correlations) to investigate the discriminating power of items.

For the analysis, we used the R platform (R Core Team, 2021). For the computation of Cronbach's alpha, item discrimination indices and EFA we used the psych package (Revelle, 2021). The reliability analysis and EFA were conducted only the respondents who gave responses to all items of each scale. To identify the number of factors to extract we used Parallel Analysis and cross-checked our findings with the Velicer MAP and Very Simple Structure techniques (Revelle & Rocklin, 1979).

The dataset for each of the four scales was analysed independently and the results are presented separately.

Interviews

The interview schedule was designed using cognitive testing. Cognitive testing is '*an evidence-based, qualitative method specifically designed to investigate whether a survey question—whether attitudinal, behavioral, or factual in nature—fulfils its intended purpose*' (Willis & Artino, 2013,p.X). During cognitive testing interviewers are empirically trying to understand the mental process through which individuals process and respond to items (Willis, 2009).

Tourangeau's (1984) 4-stage cognitive model was followed to develop the interview schedule, which includes:

1. Comprehension; 2. Retrieval of information; 3. Judgment or estimation; and 4. Selection of a response to the question. Additional questions were also asked relating to the layout, navigation and structure of the questionnaire, and its functionality. (The short timeframe for the completion of this study, inhibited us from conducting the interviews after the statistical analysis was completed and follow up on items identified problematic based on that analysis.)

Before the interviews, interviewees took part in a one-hour training session led by the project lead. The aim of the training was to support a consistent way to interviewing. Amongst other, during the training, team members discussed the interview schedule in a systematic way - resolved comprehension challenges, discussed the ordering the questions and the approach to interviewing, identified priority questions, and agreed focus and timings.

The scope of the study allowed for a total of final interviews. To identify the potential sample, the final question of the questionnaire asked respondents for their email address if they consented to being part in the interviews, who were then contacted by the research team.

Interviews lasted 30 minutes and were conducted online using Teams. Priority was given to the cognitive testing questions. During the interviews, the questionnaire items were shown on screen and read out by the interviewers. A hybrid method of interviewing was used: speak-aloud and probing.

With interviewees permission, all interviews were audio-recorded. The interviewers made notes throughout too. Final notes were typed and the interviewer manually coded the data and identified key themes. A summary of the student data was sent to the other two members of the research team, who were using the same interview schedule with parents and teachers who filled in the relevant inclusion questionnaire

for their 'respondent group'. When summaries of the data for all the respondent groups were completed, a group analysis session took place. During the analysis session the research team familiarised itself with all the data and further analysed the data further for each respondent group and across groups. Because each interviewer had been allocated a specific respondent group, the team analysis provided differing perspectives and the ability to prompt each other to critically reflect upon the analysis completed by individual interviewees and revisions to the original themes to take place, which, in turn, supported the validity of the results.

The sample

The study was conducted between February to September 2022. Overall, 165 teachers from 9 schools in the Plymouth area took part in the study. Out of the 165 completed questionnaires, circa 40% were working in schools with more than 1200 students, one third was working in schools with 901-1200 students, around one fourth was working in schools with 601-900 students and the rest were working in smaller schools.

Approximately 20% of the teachers said that they have 10-20 students in their class(es) who have a special educational need and/or disability (SEND). Around three quarters said that they have less than 10 students with a special educational need and/or disability (SEND).

Almost half of the responders were Subject Lead, one sixth of the responders were Assistant Headteacher / Principal, around a tenth were Head of Year and the rest had other roles.

As regards the interviews, out of the total of five interviewees, three were teachers, one was teacher and also held a leadership role whilst one was a headteacher with no teaching responsibilities.

The procedure

All 19 secondary schools in the city of Plymouth were invited to take part in the study. At the outset, the research team briefly introduced the study in one of the regular Headteachers' meetings convened by the Plymouth Education Board (PEB), part of Plymouth Council. Schools interested to find out more details about the study were invited to a separate meeting with the research team. During that meeting, amongst others, the research team provided detailed information about the study and answered questions. As a follow up, schools received written notes from the meeting and, at that stage, were asked to formally express their interest in taking part in the project. A total of 10 secondary schools, 9 secondary schools and one alternative provision voluntarily agreed to take part in the project.

After schools' self-selection was completed, an information pack was sent to them. Relevant information was passed to parents by the schools. Parents were asked to provide their consent before completing the questionnaire and were also informed that they could withdraw from the study at any point. Withdrawal from the study as a whole was possible until July 2022, when the analysis phase began.

Before the study began, ethical approval by the ethics panels by Plymouth Marjon University's ethics panel was sought.

Section 3 - Results

Attitudes to Inclusion Scale (AIS)

Descriptive statistics

Some of the items seem to demonstrate less variance than others. For example, item 2 “I believe that inclusion is beneficial to all students socially.” has limited variance as there are too few responses between ‘1’ and ‘5’ and almost all of the responses are ‘6’ and ‘7’ (6. Moderately agree, 7. Strongly agree). Similar results are also observed for item 7 “I believe that including students with a range of abilities will make me a better teacher.” and for item 8 “I am happy to have students who need assistance with their daily activities included in my classrooms.”.

The rest of the items demonstrate much higher variance and respondents utilized the whole range of the scale. For example, item 1 “I believe that all students, regardless of their ability, should be taught in regular classrooms.” received mixed responses.

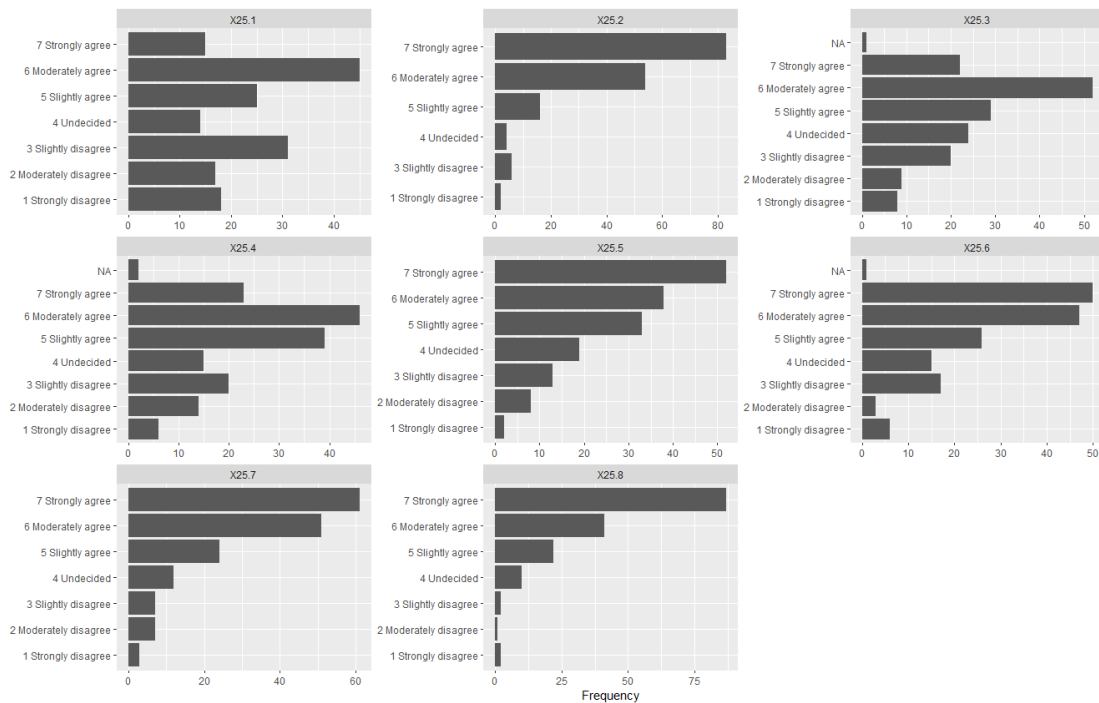


Figure 2. Frequency distributions per response per item

Item inter-correlations

We investigated the correlation matrix between the eight items of the scale. We found that there are both larger and smaller correlations between the items (Pearson correlations were used but Polychoric correlations were similar). We did not identify items with near-zero correlations with the rest of the group of items.

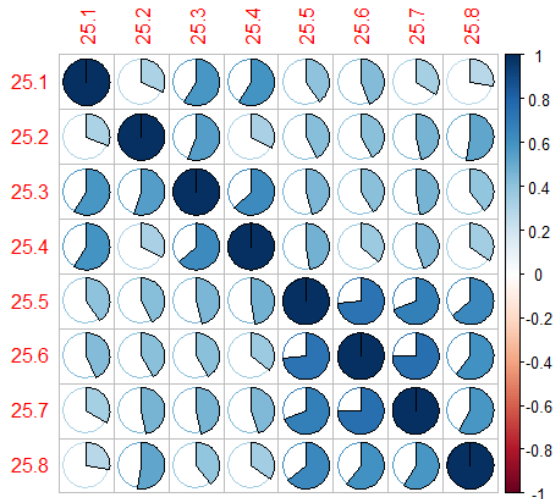


Figure 3. Correlogram of polychoric correlations between the 8 items of the scale. Larger proportion of shaded pie-charts (and darker shade) represent larger correlations.

The factor structure

Parallel Analysis using the Polychoric correlation matrix was used in order to identify the most appropriate number of factors to be extracted. Two or three factors were the best solution; a visual inspection of the Scree Plot reveals that two factors were the most reasonable solution (the third factor had an eigenvector close to zero). VSS complexity 2 achieves a maximum of 0.73 with 2 factors and the Velicer MAP achieved a minimum of 0.07 with 2 factors.

Two-factor solution: A two-factor solution suggests that items 1, 3, 4 belong to the 'Belief' factor and the other items belong to the 'Feelings' factor. Item 2 (I believe that inclusion is beneficial to all students socially) does not seem to conform to our expectations because it should load more clearly on the 'Belief' factor. However, its loading is low on both factors (0.379 and 0.282), so the item was removed from further analysis.

Loadings:

	MR1	MR2
25.1		0.725
25.2	0.379	0.281
25.3		0.849
25.4		0.759
25.5	0.806	
25.6	0.888	
25.7	0.821	
25.8	0.757	

	MR1	MR2
SS loadings	2.830	1.907
Proportion Var	0.354	0.238
Cumulative Var	0.354	0.592

The two-factor analysis was re-rerun. The main findings are shown below, but are also detailed in Appendix 1.

Loadings:

	MR1	MR2
25.1		0.752
25.3		0.763
25.4		0.811
25.5	0.813	
25.6	0.901	
25.7	0.819	
25.8	0.728	

	MR1	MR2
SS loadings	2.677	1.812
Proportion Var	0.382	0.259
Cumulative Var	0.382	0.641

According to past research, we expected items 1-4 to load on a 'Beliefs' factor and the last four items to load on a 'Feelings' factor. The factors were recovered as suggested by past research, although item 2 had to be removed from the analysis. Interestingly, the two factors are correlated ($r=0.62$) and this is significant as there seems to be a relationship between beliefs and feelings. However, item 7, which now has been reworded to read 'I believe...' instead of 'I am pleased', now loads on the 'Feelings' factor. This reduces the face validity of the factor structure, as one might expect an item starting with 'I believe' to load on the 'Believe' factor, as the other 'I believe' items do.

Factor Analysis

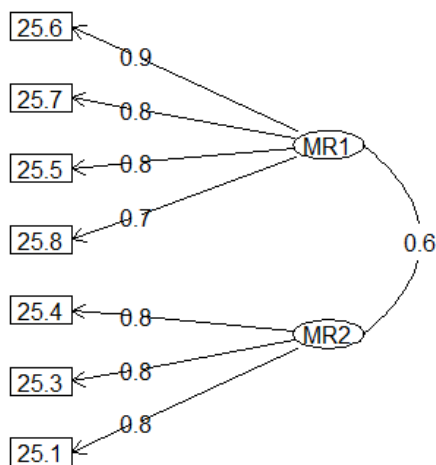


Figure 4. The factor structure of the scale.

It is important to note that other rotations (e.g., an orthogonal varimax rotation) yield the same structure with similar loadings. This suggests that our decision to use the oblimin rotation has not affected our findings.

Factor 1, Beliefs:

1. I believe that all students regardless of their ability should be taught in regular classrooms.
3. I believe that inclusion benefits all students academically.
4. I believe that all student can learn in inclusive classrooms if their teachers are willing to adapt the curriculum.

Factor 2, Feelings:

5. I am pleased that I have the opportunity to teach students with lower academic ability alongside other students in my class
6. am excited to teach students with a range of abilities in my class.
7. I believe that including students with a range of abilities will make me a better teacher.
8. I am happy to have students who need assistance with their daily activities included in my classrooms.

In addition to removing item 2 from the analysis, 7 could be removed to increase the face validity of the factor structure. This would leave us with two factors, each with three items, which is adequate for all practical intents and purposes.

Reliability and Discriminating power of items

For each of the factors, we computed Cronbach's alpha. Before running the analysis, we removed item 7 from the analysis, as suggested above.

The Cronbach's alpha for the first factor ('Beliefs'), based on polychoric correlation matrix, was 0.82. The corrected item-total correlation for the items of the factor ranged from 0.72 to 0.76 which is satisfactory for all intents and purposes, suggesting that all items contribute significantly to the measurement exercise. The average item inter-correlation is 0.60, well within the recommendations of Clark & Watson (1995), which suggests that the items tap on the same construct, as intended.

The Cronbach's alpha for the second factor ('Feelings'), based on polychoric correlation matrix, was 0.85. The corrected item-total correlation for the items of the factor ranged from 0.72 to 0.84 which is satisfactory for all intents and purposes. The average inter-item correlation was high, at 0.66, which suggests that the items tap on the same construct, as intended.

These are very satisfactory values for Cronbach's alpha, for scales with a small number of items, as the value of the index tends to shrink while the number of items in the factor is reduced (Cortina, 1993; Nunnally, 1994; Streiner, 2003). Overall, the scales seem to have very satisfactory reliability indices and high inter-item correlations.

Triangulating qualitative and quantitative data

As it has already been indicated by the quantitative analysis, some of the items of the scale may have certain shortcomings, probably caused by the number of options used in the response scale: for a scale of seven 'steps' (1-7) or six steps (1-6) some teachers would prefer to have the option for a 'neutral' or 'N/A' option, while others felt that not having this pushed them towards a 'truer' answer (respondents looked at options, wanting a neutral answer, but then had to decide between binary options and then to what extent)

Regarding the length of time working in the school: it is not clear how this may have affected answers in any way. The question about 'how confident do you feel about your answers' has been helpful with this. Breadth of previous teaching experience may also contribute to respondents' perspectives. A free-text box at the end for 'other comments' could be useful, or further questions about how confident they feel about all answers.

During interviews, some teachers suggested that there might be some ambiguity of wording. For example, it was not clear what might be meant by terms such as 'all students', 'regular classroom' etc. For example, see item 2 (I believe that inclusion is beneficial to all students socially). It was suggested that adding additional explanation in parentheses could help. The quantitative analysis has indeed shown that this particular item should be removed from further analysis as it loaded on two different factors equally, however, this might or might not be related to the wording problems because other items with similar issues had very pleasant psychometric characteristics.

It is important, however, to note that if any changes are attempted, the instrument would need to be re-piloted and re-evaluated for its psychometric characteristics.

Recommendations

Some of the items of the AIS are not needed and might be removed from future administrations of the instrument (Item 2 and Item 7). Removing these items would shorten the scale a little bit, without significant loss of information, and would also improve the face validity of the instrument.

Item 7 seems to have been re-worded compared to earlier versions of the instrument (from past research). The item now loads on the 'wrong' factor. It needs to be decided whether the item would be removed or re-worded back to its original wording.

For purposes of further analysis, we recommend to aggregate two 'scores' for each teacher, based on the two factors revealed: 'Beliefs' and 'Feelings'. These scores could be used to investigate differences between groups with different demographic characteristics or between schools etc.

The scale has very satisfactory psychometric characteristics and our results have largely corroborated the finding of past research. Further changes are not required. However, if changes are attempted, the instrument would need to be re-piloted and re-evaluated for its psychometric characteristics.

Intention to Teach in Inclusive Classrooms (ITIC)

Descriptive statistics

Some of the items do not have the same variance as others. For example, item 2 “Consult with the parents/carers of a student who is struggling in your class.” has a very restricted range of values with most of the responses being between 5 (Somewhat likely) to 7 (Extremely likely). The same holds true for item 3 “Consult with your colleagues to identify possible ways you can assist a struggling student in your class.” and for item 5 “Consult with a student who is displaying challenging behaviors to find out better ways to work with them.”. Some other items demonstrate responses which cover the whole range of the scale but the distribution is skewed towards higher values; for example, see item 7 “Change the assessment tasks to suit the learning profile of a student who is struggling (e.g. provide more time to complete the task or modify test questions).” Finally, the other items demonstrate distributions which make use of the whole scale and the distributions are less skewed.

It is generally preferable not to have too many items in one scale with low variance, as the purpose of a typical survey is to identify trends and patterns. If most teachers give the same or very similar response to questions, our aim is not achieved.

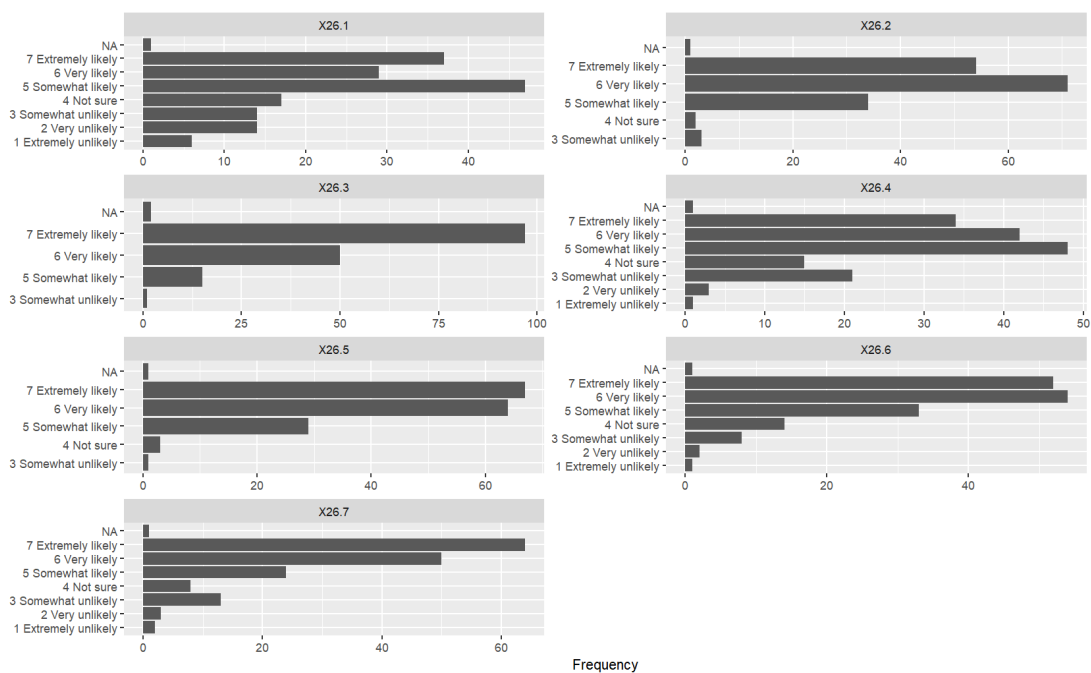


Figure 5. Frequency distributions per response per item

Item inter-correlations

We investigated the correlation matrix between the seven items of the scale. We found that there are both larger and smaller correlations between the items (Pearson correlations were used but Polychoric correlations were similar). We did not identify items with near-zero correlations with the rest of the group of items.

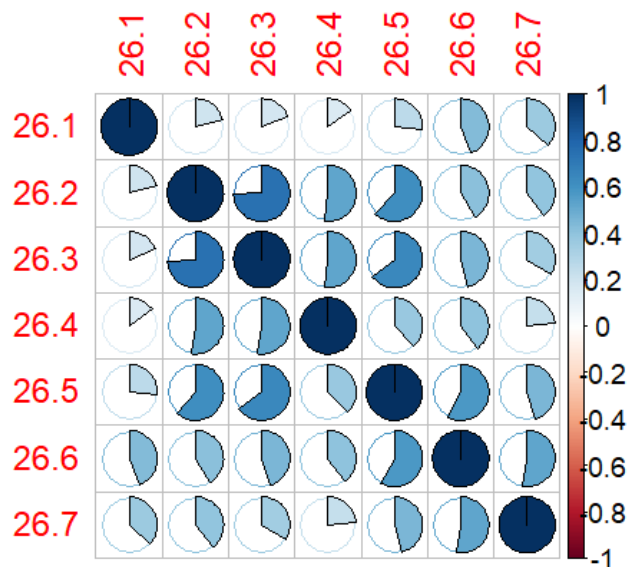


Figure 6. Correlogram of polychoric correlations between the items of the scale. Larger proportion of shaded pie-charts (and darker shade) represent larger correlations.

The factor structure

Parallel Analysis using the Polychoric correlation matrix was used in order to identify the most appropriate number of factors to be extracted. Two factors were the best solution of Parallel Analysis; a visual inspection of the Scree Plot reveals that one or two factors were the most reasonable solution (the second and third factors had an eigenvector close to zero). VSS complexity 2 achieves a maximum of 0.73 with 2 factors and the Velicer MAP achieved a minimum of 0.07 with a single factor.

Two-factor solution: A two-factor solution recovers the same structure as the one expected by the literature: one factor is the 'Intentions for curriculum change' (items 1, 6, 7) and one factor is the 'Intentions to consult' (items 2, 3, 4). There is a correlation of 0.6 between the two factors. However, there seems to be an issue with item 5: its loading is low on both factors (0.513 and 0.342), so the item was removed from further analysis.

Loadings:

	MR1	MR2
26.1	-0.152	0.630
26.2	0.846	
26.3	0.912	
26.4	0.549	
26.5	0.513	0.342
26.6		0.781
26.7		0.634

	MR1	MR2
SS loadings	2.141	1.531
Proportion Var	0.306	0.219
Cumulative Var	0.306	0.525

The two-factor analysis was re-rerun. The main findings are shown below, but are also detailed in Appendix 2.

Loadings:

	MR1	MR2
26.1	-0.150	0.649
26.2	0.842	
26.3	0.875	
26.4	0.573	
26.6		0.748
26.7		0.627

	MR1	MR2
SS loadings	1.838	1.379
Proportion Var	0.306	0.230
Cumulative Var	0.306	0.536

Our findings seem to agree with past research, with each of the items loading clearly on a single factor, as expected.

Factor Analysis

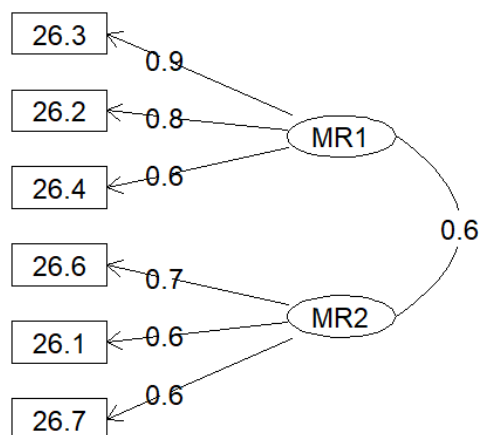


Figure 7. The factor structure of the scale.

Factor 'Intentions for curriculum change':

1. Change the curriculum to meet the learning needs of a student with learning difficulty enrolled in your class.
6. Include students with severe disabilities in a range of social activities in your class.
7. Change the assessment tasks to suit the learning profile of a student who is struggling (e.g., providing longer time to complete the task or modifying test questions).

Factor 'Intentions to consult':

2. Consult with the parents of a student who is struggling in your class.
3. Consult with your colleagues to identify possible ways you can assist a struggling student in your class.
4. Undertake a professional development program so you can teach students with diverse learning needs well.

Note: It is important to note that other rotations (e.g., an orthogonal varimax rotation) yield the same structure with similar loadings. This suggests that our decision to use the oblimin rotation has not affected our findings.

Reliability and Discriminating power of items

For each of the factors, we computed Cronbach's alpha. The Cronbach's alpha for the first factor ('Intentions for curriculum change'), based on polychoric correlation matrix, was 0.71. The corrected item-total correlation for the items of the factor ranged from 0.55 to 0.70 which is satisfactory for all intents and purposes, suggesting that all items contribute significantly to the measurement exercise. The average item inter-correlation is 0.44, marginally within the recommendations of Clark & Watson (1995), which is still satisfactory but also suggests that the items tap on the same construct, as intended.

The Cronbach's alpha for the second factor ('Intentions to consult'), based on polychoric correlation matrix, was 0.81. The corrected item-total correlation for the items of the factor ranged from 0.60 to 0.82 which is satisfactory for all intents and purposes. The average inter-item correlation was high, at 0.59, which suggests that the items tap on the same construct, as intended.

These are very satisfactory values for Cronbach's alpha, for scales with only three items, as the value of the index tends to shrink while the number of items in the factor is reduced (Cortina, 1993; Nunnally, 1994; Streiner, 2003). Overall, the scales seem to have very satisfactory reliability indices and high inter-item correlations.

Triangulating qualitative and quantitative data

During interviews, some teachers suggested that there might be some ambiguity of wording but there were no specific examples of items of this scale which may have been too ambiguous.

As it has already been indicated elsewhere, some of the items of the scale may have certain shortcomings, probably caused by the number of options used in the response scale: for a scale of seven 'steps' (1-7) some teachers would prefer to have the option for a 'neutral' or 'N/A' option, while others felt that not having this pushed them towards a 'truer' answer (respondents looked at options, wanting a neutral answer, but then had to decide between binary options and then to what extent)

Regarding the length of time working in the school: it is not clear how this may have affected answers in any way. The question about 'how confident do you feel about your answers' has been helpful with this. Breadth of previous teaching experience may also contribute to respondents' perspectives. A free-text box at the end for 'other comments' could be useful, or further questions about how confident they feel about all answers.

It is important, however, to note that if any changes are attempted, the instrument would need to be re-piloted and re-evaluated for its psychometric characteristics.

Recommendations

Item 5 of the ITIC scale is not needed and might be removed from future administrations of the instrument. Removing the item would shorten the scale a little bit, without significant loss of information.

For purposes of further analysis, we recommend to aggregate two 'scores' for each teacher, based on the two factors revealed: 'Intentions for curriculum change and 'Intentions to consult''. These scores could be used to investigate differences between groups with different demographic characteristics or between schools etc.

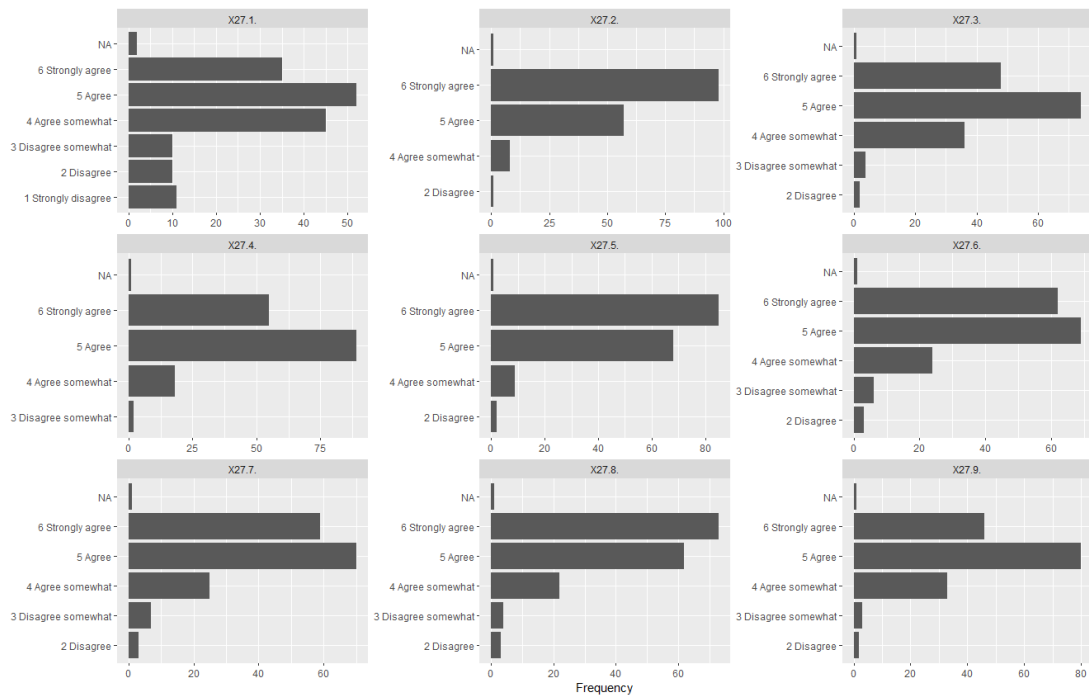
The scale has very satisfactory psychometric characteristics and our results have largely corroborated the finding of past research. Further changes are not required. However, if changes are attempted, the instrument would need to be re-piloted and re-evaluated for its psychometric characteristics.

Teaching Efficacy to implement Inclusive Practices- TEIP

Descriptive statistics

Some of the items do not have the same variance as others. For example, item 2 “I am able to provide an alternate explanation or example when students are confused.” has a very restricted range of values with most of the responses being between 5 (Agree) to 6 (Strongly Agree). Other items demonstrate distributions which make use of the whole scale and the distributions are less skewed such as item 18 “I am confident in informing others who know little about laws and policies relating to the inclusion of students who have SEND.”.

It is generally preferable not to have too many items in one scale with low variance, as the purpose of a typical survey is to identify trends and patterns. If most teachers give the same or very similar response to questions, our aim is not achieved.



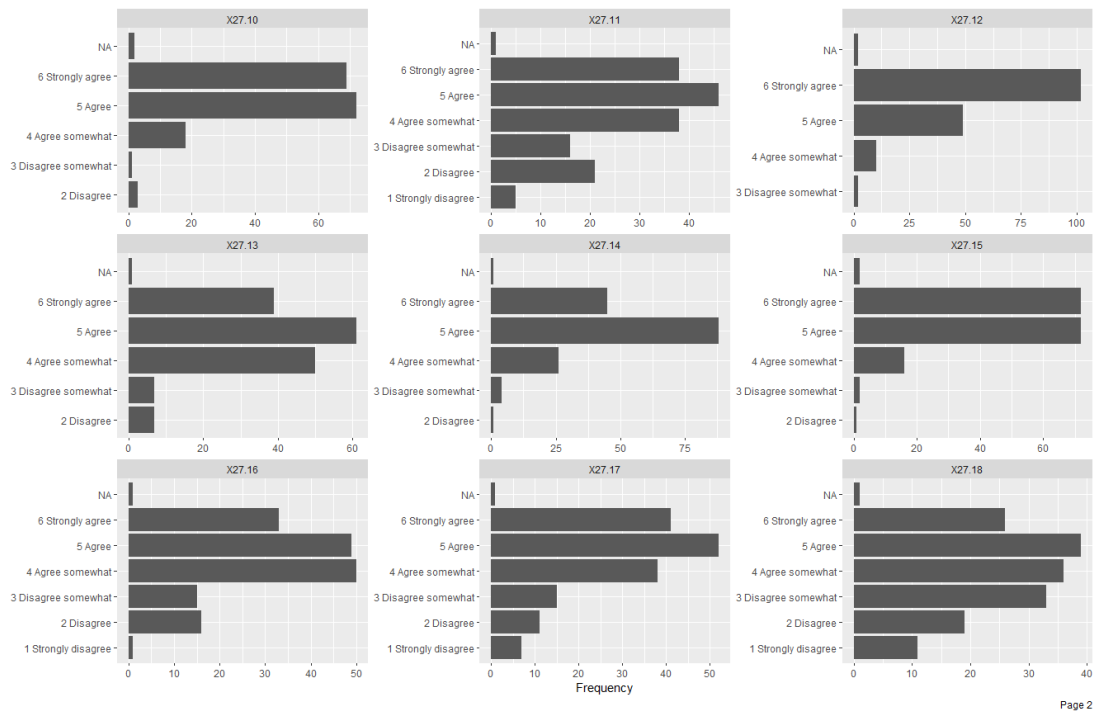


Figure 8. Frequency distributions per response per item

Item inter-correlations

We investigated the correlation matrix between the eighteen items of the scale. We found that there are both larger and smaller correlations between the items (Pearson correlations were used but Polychoric correlations were similar). We did not identify items with near-zero correlations with the rest of the group of items. However, some of the items (e.g. item 1) seem to have somewhat lower correlations, overall, and they may not load properly on any of the factors when we attempt an EFA.

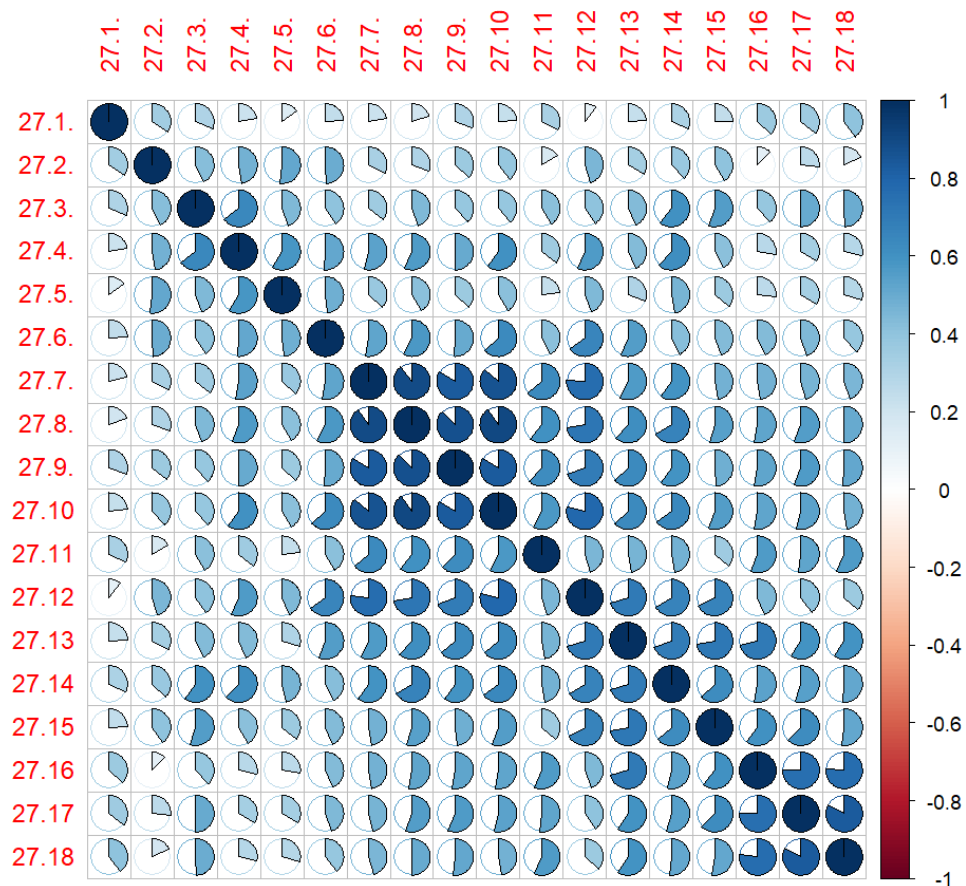


Figure 9. Correlogram of polychoric correlations between the items of the scale. Larger proportion of shaded pie-charts (and darker shade) represent larger correlations.

The factor structure

Parallel Analysis using the Polychoric correlation matrix was used in order to identify the most appropriate number of factors to be extracted. Between three and four factors were the best solution of Parallel Analysis, but the fourth factor had a very low eigenvalue. A visual inspection of the Scree Plot reveals that three factors were the most reasonable solution (other factors had an eigenvector close to zero). VSS complexity 2 achieves a maximum of 0.78 with 2 factors and the Velicer MAP achieved a minimum of 0.04 with three factors.

We decided to attempt an EFA with three factors.

Three-factor solution: A three-factor solution recovers somewhat similar but NOT exactly the same structure as the one expected by the literature.

Loadings:

	MR1	MR2	MR3
27.1.	-0.104	0.394	0.196
27.2.		0.718	
27.3.	-0.137	0.355	0.637
27.4.	0.250		0.671
27.5.		0.647	
27.6.	0.369		0.378
27.7.	0.985		
27.8.	0.917		
27.9.	0.821	0.137	
27.10	0.900		
27.11	0.463	0.377	
27.12	0.698		0.295
27.13	0.345	0.445	0.163
27.14	0.280	0.300	0.400
27.15	0.131	0.439	0.348
27.16	0.124	0.847	
27.17		0.837	
27.18		0.919	

	MR1	MR2	MR3
SS loadings	4.462	3.211	2.398
Proportion Var	0.248	0.178	0.133
Cumulative Var	0.248	0.426	0.559

The first observation is that some items need to be removed from the analysis. For example, item 1 seems to have similarly low loadings on multiple factors. Item 6 also loads on both the first and the third factor, so it is not convenient to keep it in the analysis. For the same reason, we are motivated to remove items 11, 13, 14 and 15.

Note: a two-factor and a four-factor solution did not improve the situation as a large number of items still loaded similarly on multiple factors.

We re-run the EFA, after removing items 1, 6, 11, 13, 14, 15.

Loadings:

	MR1	MR2	MR3
27.2.		0.645	
27.3.	-0.130	0.328	0.660
27.4.	0.216		0.724
27.5.		0.669	
27.7.	0.977		
27.8.	0.914		
27.9.	0.833	0.126	
27.10	0.911		
27.12	0.706		0.236
27.16	0.183	0.761	
27.17		0.854	
27.18		0.937	

	MR1	MR2	MR3
SS loadings	3.916	2.333	1.899
Proportion Var	0.326	0.194	0.158
Cumulative Var	0.326	0.521	0.679

The results suggest that the items cluster in three factors:

Factor 1 - Efficacy in inclusive instruction:

2. I am able to provide an alternate explanation or example when students are confused.
3. I am confident in designing learning tasks so that the individual needs of students with disabilities are accommodated.
4. I can accurately gauge what progress students have made in a taught area.
5. I can provide appropriate challenges for very capable students

It is reminded that past research has also suggested an 'Efficacy in inclusive instruction' factor with items 1-6. In a sense, our results recover the same factor, although some of the items do not seem to conform to our expectation.

Factor 2 - Efficacy in managing behavior:

7. I am confident in my ability to prevent disruptive behaviour in the classroom before it occurs
8. I can control disruptive behaviour in the classroom
9. I am able to calm a student who is disruptive or noisy
10. I am able to get children to follow classroom rules
11. I am confident when dealing with students who are physically aggressive
12. I can make my expectations clear about student behaviour

It is reminded that past research has also suggested an 'Efficacy in managing behavior' factor with items 7-12. Thus, our results recover perfectly the same factor.

Factor 3 - Efficacy in collaboration:

16. I am confident in my ability to get parents/carers involved in the school activities of their children who have SEND
17. I can collaborate with other professionals (e.g. SENDCo, Educational Psychologist, Speech and Language Therapist) in designing educational plans for students who have SEND
18. I am confident in informing others who know little about laws and policies relating to the inclusion of students who have SEND

It is reminded that past research has also suggested an 'Efficacy in collaboration' factor with items 13-18. In a sense, our results recover the same factor, although some of the items do not seem to conform to our expectation.

Our findings seem to agree with past research, with each of the items loading clearly on a single factor, as expected.

Factor Analysis

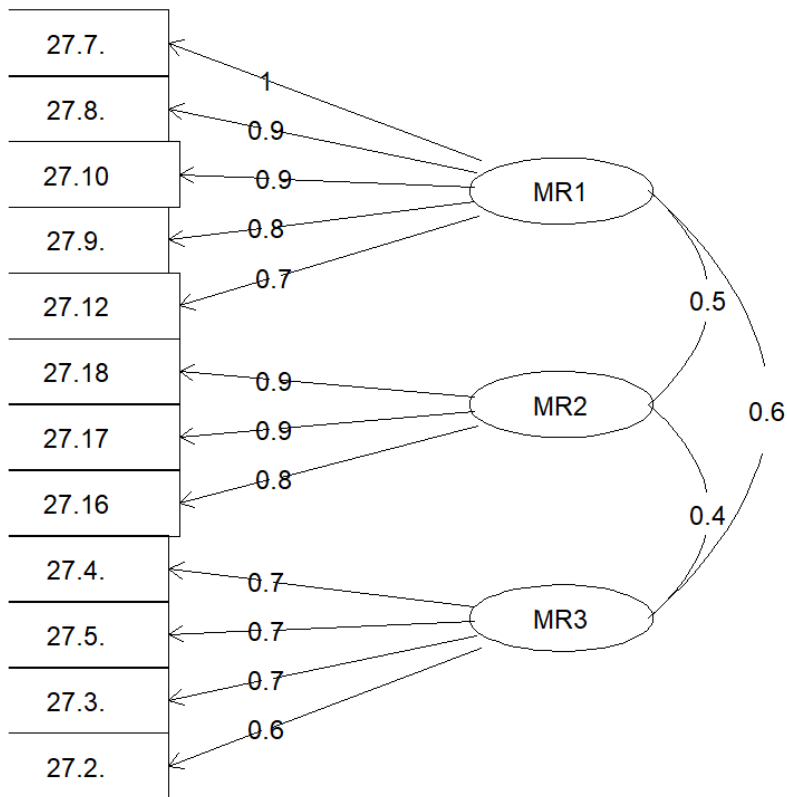


Figure 10. The factor structure of the scale.

Interestingly, the three factors are correlated significantly. This is reasonable and expected, in the sense that the dimensions (factors) uncovered indeed seem to be related, at least at face value. Past research attempted varimax rotations. It is important to note that other rotations (e.g., an orthogonal varimax rotation) yield very similar structures with similar loadings. This suggests that our decision to use the oblimin rotation has not affected our findings.

Reliability and Discriminating power of items

For each of the factors, we computed Cronbach's alpha. The Cronbach's alpha for the first factor ('Efficacy in inclusive instruction'), based on polychoric correlation matrix, was 0.81. The corrected item-total correlation for the items of the factor ranged from 0.63 to 0.80 which is satisfactory for all intents and purposes, suggesting that all items contribute significantly to the measurement exercise. The average item inter-correlation is 0.52, within the recommendations of Clark & Watson (1995), which is still satisfactory but also suggests that the items tap on the same construct, as intended.

The Cronbach's alpha for the second factor ('Efficacy in managing behavior'), based on polychoric correlation matrix, was 0.94. The corrected item-total correlation for the items of the factor ranged from 0.65 to 0.95 which is satisfactory for all intents and

purposes. The average inter-item correlation was very high, at 0.74, which suggests that the items tap on the same construct, as intended.

The Cronbach's alpha for the third factor ('Efficacy in collaboration'), based on polychoric correlation matrix, was 0.92. The corrected item-total correlation for the items of the factor ranged from 0.82 to 0.90 which is satisfactory, suggesting that all items contribute significantly to the measurement exercise. The average item inter-correlation is 0.79, is well within the recommendations of Clark & Watson (1995), which is satisfactory but also suggests that the items tap on the same construct, as intended.

These are very satisfactory values for Cronbach's alpha, even for scales with only three items, as the value of the index tends to shrink while the number of items in the factor is reduced (Cortina, 1993; Nunnally, 1994; Streiner, 2003). Overall, the scales seem to have very satisfactory reliability indices and high inter-item correlations.

Triangulating qualitative and quantitative data

During interviews, some teachers suggested that there might be some ambiguity of wording. It is not clear if there were ambiguous items in this scale, as the qualitative data did not have specific information.

As it has already been indicated elsewhere, some of the items of the scale may have certain shortcomings, probably caused by the number of options used in the response scale: for a scale of seven 'steps' (1-7) some teachers would prefer to have the option for a 'neutral' or 'N/A' option, while others felt that not having this pushed them towards a 'truer' answer (respondents looked at options, wanting a neutral answer, but then had to decide between binary options and then to what extent)

Regarding the length of time working in the school: it is not clear how this may have affected answers in any way. The question about 'how confident do you feel about your answers' has been helpful with this. Breadth of previous teaching experience may also contribute to respondents' perspectives. A free-text box at the end for 'other comments' could be useful, or further questions about how confident they feel about all answers.

It is important, however, to note that if any changes are attempted, the instrument would need to be re-piloted and re-evaluated for its psychometric characteristics.

Recommendations

From the analysis above, we recommend the following.

A number of items (Items 1, 6, 11, 13, 14, 15) need to be removed from the analysis because they load on more than one factors or they have very low loadings. Removing the items would shorten the scale substantially, but we are still able to successfully perform an Exploratory Factor Analysis to reveal the structure of the instrument.

The analysis revealed three reliable and meaningful factors: 'Efficacy in inclusive instruction', 'Efficacy in managing behavior', 'Efficacy in collaboration'. To some

degree these results corroborate those of past research, although not perfectly. This is something that needs to be theoretically studied.

The 'Efficacy in collaboration' factor included only three items; this is the minimum number of items for a factor we would typically require.

For purposes of further analysis, we recommend to aggregate three 'scores' for each teacher, based on the three factors revealed. These scores could be used to investigate differences between groups with different demographic characteristics etc.

The scale (after dropping a number of items) has very satisfactory psychometric characteristics and our results have corroborated to some degree the findings of past research. Further changes in the wording etc are neither required nor suggested. However, if changes are attempted, the instrument would need to be re-piloted and re-evaluated for its psychometric characteristics.

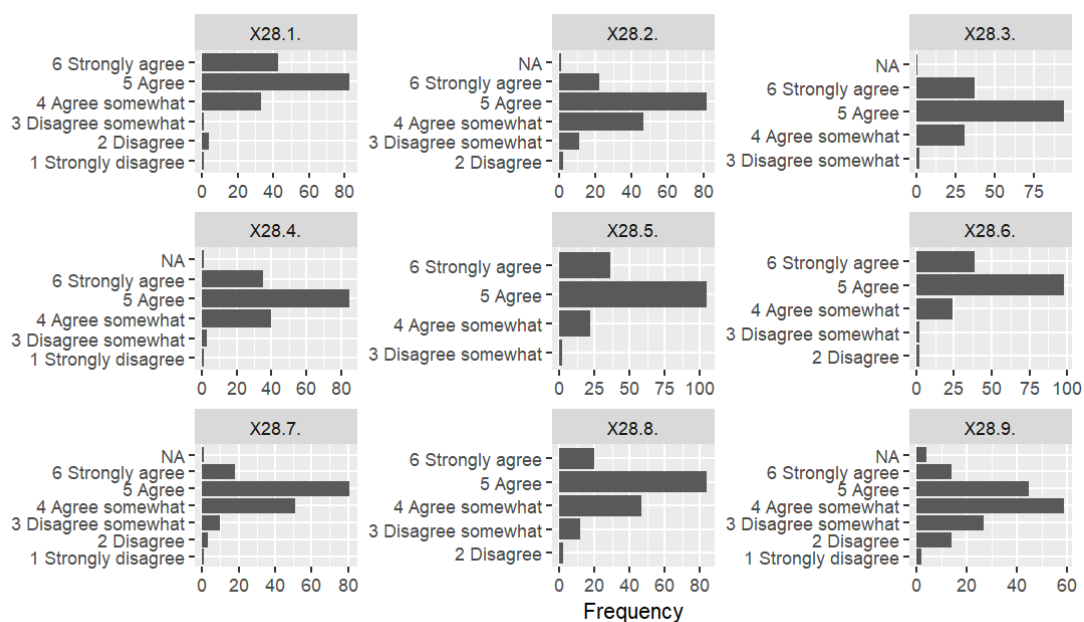
Teaching Efficacy to implement Inclusive Practices Collective Scale – (TEIP-C)

Descriptive statistics

As was the case with the TEIP scale, some of the items do not have the same variance as others. For example, item 6 “Teachers in my school can provide appropriate challenges for very capable students.” has a very restricted range of values with most of the responses being between 5 (Agree) to 6 (Strongly Agree). Similarly, item 5 mainly attracted ‘Agree’ responses.

Other items demonstrate distributions which make use of the whole scale and the distributions are less skewed such as item 9 “Teachers in my school can get parents/carers of children who are frequently excluded involved in school activities.”

It is generally preferable not to have too many items in one scale with low variance, as the purpose of a typical survey is to identify trends and patterns. If most teachers give the same or very similar response to questions, our aim is not achieved.



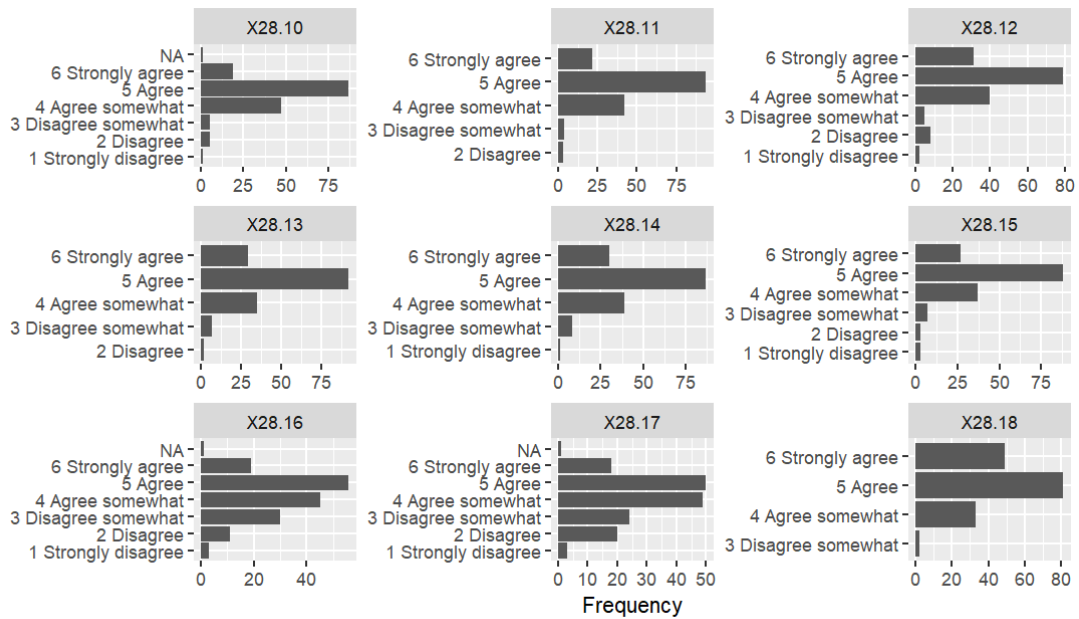


Figure 11. Frequency distributions per response per item

Item inter-correlations

We investigated the correlation matrix between the eighteen items of the scale. We found that there are few large and many smaller correlations between the items (Pearson correlations were used but Polychoric correlations were similar). We did not identify items with near-zero correlations with the rest of the group of items.

Items 7 and 8 ("Teachers in my school can prevent disruptive behaviour in the classroom before it occurs" and "Teachers in my school can control disruptive behaviour in the classroom") have a particularly high correlation ($r=0.82$) and it is likely that one of the two may be redundant.

Another pair of items with high correlation ($r=0.72$) is items 3 and 4 ("Teachers in my school can make parents/carers feel comfortable coming to school" and "Teachers in my school can assist families in helping their children do well in school").

Beyond some pairs of items with high correlations, however, it seems that the correlations between the items are rather low (most correlations between 0.3-0.5). This could mean that an EFA may not yield stable factors.

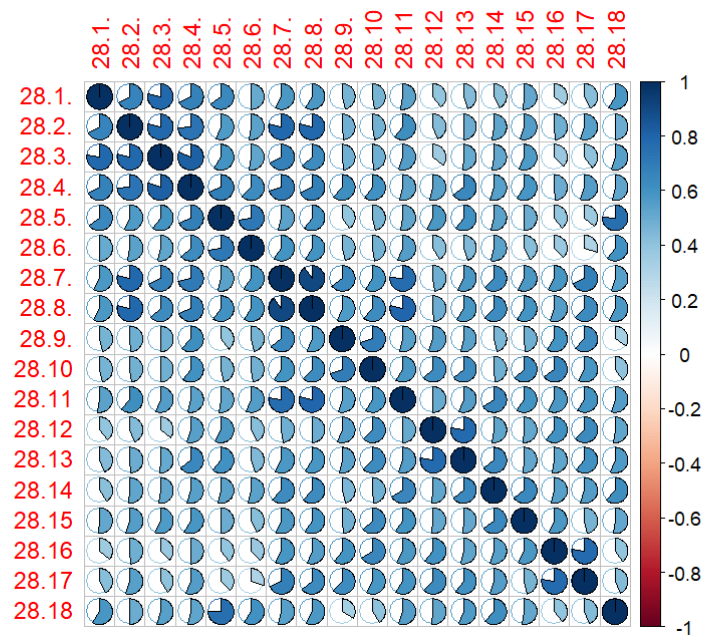


Figure 12. Correlogram of polychoric correlations between the items of the scale. Larger proportion of shaded pie-charts (and darker shade) represent larger correlations.

The factor structure

Parallel Analysis using the Polychoric correlation matrix was used in order to identify the most appropriate number of factors to be extracted. Between two and four factors were the best solutions of Parallel Analysis, but the third and fourth factors had a very low eigenvalue. A visual inspection of the Scree Plot reveals that two to three factors were the most reasonable solution (the third had a very small eigenvalue). VSS complexity 2 achieves a maximum of 0.77 with 2 factors and the Velicer MAP achieved a minimum of 0.04 with four factors, but the reduction in map value between the third and fourth factor is very small.

We decided to first attempt an EFA with two factors. Since the TEIP-C scale is very similar to the one used for the TEIP analysis, we also attempted a three-factor model.

Two-factor solution: A two-factor solution was run on the data.

Loadings:

	MR1	MR2
28.1.	0.870	
28.2.	0.748	0.114
28.3.	0.956	-0.147
28.4.	0.759	0.122
28.5.	0.805	
28.6.	0.704	
28.7.	0.565	0.382
28.8.	0.590	0.355
28.9.	0.117	0.673

28.10	0.128	0.712
28.11	0.418	0.460
28.12	0.752	
28.13	0.230	0.620
28.14	0.438	0.381
28.15	0.445	0.358
28.16	-0.118	0.923
28.17	0.868	
28.18	0.689	

	MR1	MR2
SS loadings	5.756	4.335
Proportion Var	0.320	0.241
Cumulative Var	0.320	0.561

Five items needed to be removed, because they loaded similarly on two factors. The analysis was re-run and the results are presented below:

Loadings:

	MR1	MR2
28.1.	0.863	
28.2.	0.696	0.149
28.3.	0.931	-0.106
28.4.	0.751	0.165
28.5.	0.813	
28.6.	0.679	
28.9.	0.109	0.692
28.10	0.121	0.724
28.12	0.783	
28.13	0.226	0.640
28.16	-0.114	0.910
28.17	0.860	
28.18	0.671	

	MR1	MR2
SS loadings	4.327	3.668
Proportion Var	0.333	0.282
Cumulative Var	0.333	0.615

The two-factor solution yielded the following results.

Factor 1, items 9, 10, 12, 13, 16, 17, 'Efficacy in inclusive instruction and collaboration':

- 9. Teachers in my school can get parents of children who are frequently excluded involved in school activities.
- 10. Teachers in my school design learning tasks so that the individual learning needs of all students are accommodated.
- 12. Teachers in my school can collaborate with other professionals (e.g., itinerant teachers or speech pathologists) in designing educational plans for students with disabilities.

13. Teachers in my school are able to work jointly with other professionals and staff (e.g. aides, other teachers) to ensure that all students are included in the classroom.
16. Teachers in my school are confident in informing others who know little about laws and policies relating to the inclusion of students with disabilities.
17. Teachers in my school are confident in managing a situation if a student becomes physically aggressive in the classroom.

Item 17 does not seem to fit very well the rest of the items of the factor. A similar item, in the TEIP scale (see previous section), was part of the 'managing behavior' factor.

Factor 2, items 1-6 & 18, 'Efficacy in managing behavior':

1. Teachers in my school can make their expectations clear about desired student behaviour.
2. Teachers in my school are able to calm a student who is disruptive or noisy.
3. Teachers in my school can make parents feel comfortable coming to school.
4. Teachers in my school can assist families in helping their children do well in school.
5. Teachers in my school can accurately gauge what progress students have made in a taught area.
6. Teachers in my school can provide appropriate challenges for very capable students.
18. Teachers in my school are able to provide an alternate explanation or example when students are confused.

Item 18 does not seem to fit very well the rest of the items of the factor. A similar item, in the TEIP scale (see previous section), was part of the 'Efficacy in inclusive instruction' factor.

From the above, we would recommend to remove items 17 and 18, in order to have more homogeneous factors (details in Appendix 4).

Removing items 17 and 18 and rerunning the analysis yields the following results Loadings:

	MR1	MR2
28.1.	0.846	
28.2.	0.747	0.113
28.3.	0.998	-0.133
28.4.	0.761	0.184
28.5.	0.619	0.222
28.6.	0.539	0.192
28.9.	0.132	0.671
28.10		0.787
28.12		0.894
28.13	0.123	0.735
28.16		0.794

	MR1	MR2
SS loadings	3.568	3.190
Proportion Var	0.324	0.290
Cumulative Var	0.324	0.614

Factor Analysis

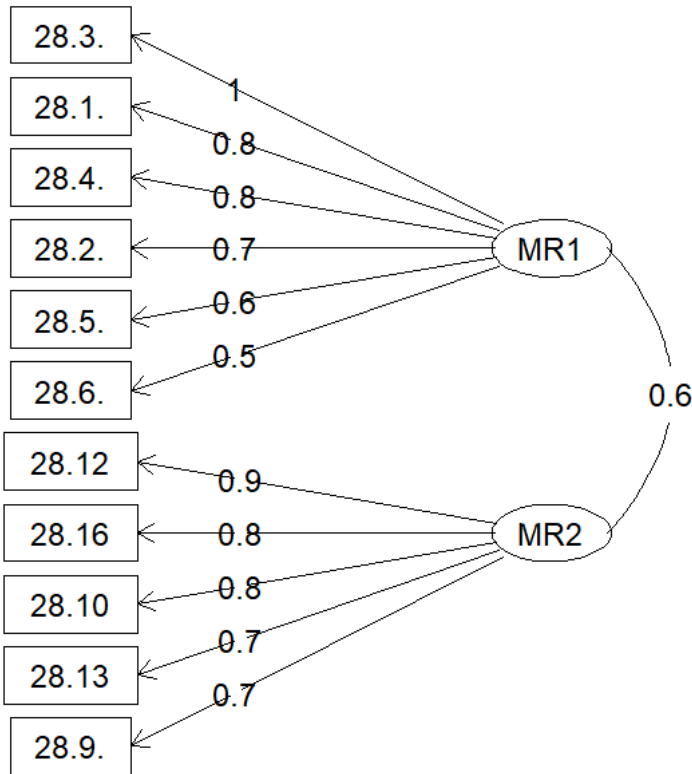


Figure 13. The factor structure of the scale.

The two factors are significantly correlated ($r=0.64$) which suggests that the teachers tend to express similar views regarding the efficiency of the teachers in managing behavior and in inclusive instruction and collaboration.

Three-factor solution: A three-factor solution yielded the following results.

A number of items load on more than one factor and will be removed from further analysis (items with red; item 6, 11, 14 and 15).

Loadings:

	MR1	MR2	MR3
28.1.	0.704		0.246
28.2.	0.863		
28.3.	0.900	-0.159	0.140
28.4.	0.670	0.123	0.178
28.5.	0.294		0.715
28.6.	0.420		0.391
28.7.	0.752	0.295	-0.102
28.8.	0.686	0.293	
28.9.	0.295	0.596	-0.115
28.10	0.182	0.664	

28.11	0.477	0.411	
28.12	-0.189	0.831	0.312
28.13	0.672	0.320	
28.14	0.282	0.399	0.257
28.15	0.374	0.351	0.159
28.16	0.824	-0.110	
28.17	0.193	0.771	-0.153
28.18	0.267	0.156	0.572

	MR1	MR2	MR3
SS loadings	4.516	3.921	1.455
Proportion Var	0.251	0.218	0.081
Cumulative Var	0.251	0.469	0.550

It is interesting that some items load on all three factors, albeit with very low loadings (and this is another reason to remove them). For example, items 14 and 15 ("Teachers in my school can get students to work together co-operatively in pairs or small groups." and "Teachers in my school can use a variety of assessment strategies in order to determine if all children in a class are learning (e.g. portfolio assessment, modified tests, performance-based assessment etc.).") have very low loadings and load similarly on all three factors.

The items highlighted in red were removed and the analysis was rerun.

Loadings:

	MR1	MR2	MR3
28.1.	0.793		0.182
28.2.	0.860		
28.3.	0.973	-0.153	
28.4.	0.726	0.165	
28.5.	0.502	0.163	0.566
28.7.	0.696	0.319	-0.185
28.8.	0.658	0.315	
28.9.	0.233	0.634	-0.174
28.10	0.165	0.691	
28.12	-0.135	0.887	0.283
28.13	0.704	0.263	
28.16	0.826	-0.148	
28.17	0.138	0.776	-0.172
28.18	0.434	0.196	0.430

	MR1	MR2	MR3
SS loadings	4.328	3.770	0.829
Proportion Var	0.309	0.269	0.059
Cumulative Var	0.309	0.578	0.638

Two more items (highlighted in red) were removed and the analysis was re-run.

Loadings:

	MR1	MR2	MR3
28.1.	0.102	0.680	0.102
28.2.		0.473	0.522
28.3.		0.948	

28.4. 0.271 0.631 0.114
 28.7. 0.128 0.860
 28.8. 0.165 0.743
 28.9. 0.560 0.212
 28.10 0.682 0.121
 28.12 0.919 -0.125
 28.13 0.755 0.197
 28.16 0.648 -0.170 0.338
 28.17 0.562 -0.143 0.463

MR1 MR2 MR3
 SS loadings 3.021 2.141 1.990
 Proportion Var 0.252 0.178 0.166
 Cumulative Var 0.252 0.430 0.596

Removing items does not lead to stable factors as new items appear to be loading on more than one factors and need to be removed. A three-factor solution does not seem to be an appropriate model for our data.

A Confirmatory Factor Analysis model was attempted to investigate the possibility for the data to support a single factor. The results were discouraging. Even after allowing for covariances for four pairs of items, fit statistics were far from satisfactory.

lavaan 0.6-10 ended normally after 44 iterations

Estimator	ML		
Optimization method	NLMINB		
Number of model parameters	39		
	Used	Total	
Number of observations		157	165

Model Test User Model:

Test statistic	413.501
Degrees of freedom	132
P-value (Chi-square)	0.000

Model Test Baseline Model:

Test statistic	1958.613
Degrees of freedom	153
P-value	0.000

User Model versus Baseline Model:

Comparative Fit Index (CFI)	0.844
Tucker-Lewis Index (TLI)	0.819

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-2801.861
-------------------------------	-----------

Loglikelihood unrestricted model (H1)	-2595.111
Akaike (AIC)	5681.722
Bayesian (BIC)	5800.916
Sample-size adjusted Bayesian (BIC)	5677.465

Root Mean Square Error of Approximation:

RMSEA	0.117
90 Percent confidence interval - lower	0.104
90 Percent confidence interval - upper	0.129
P-value RMSEA \leq 0.05	0.000

Standardized Root Mean Square Residual:

SRMR	0.075
------	-------

Reliability and Discriminating power of items

For each of the factors, we computed Cronbach's alpha. The Cronbach's alpha for the first factor ('Efficacy in managing behavior'), based on polychoric correlation matrix, was 0.92. The corrected item-total correlation for the items of the factor ranged from 0.70 to 0.88 which is satisfactory for all intents and purposes, suggesting that all items contribute significantly to the measurement exercise. The average item inter-correlation is 0.66, within the recommendations of Clark & Watson (1995), which is satisfactory but also suggests that the items tap on the same construct, as intended.

The Cronbach's alpha for the second factor ('Efficacy in inclusive instruction and collaboration'), based on polychoric correlation matrix, was 0.90. The corrected item-total correlation for the items of the factor ranged from 0.76 to 0.83 which is satisfactory for all intents and purposes. The average inter-item correlation was high, at 0.65, which suggests that the items tap on the same construct, as intended.

These are very satisfactory values for Cronbach's alpha, even for scales with only three items, as the value of the index tends to shrink while the number of items in the factor is reduced (Cortina, 1993; Nunnally, 1994; Streiner, 2003). Overall, the scales seem to have very satisfactory reliability indices and high inter-item correlations.

Analysis of interviews

Comprehension

Teachers found the questionnaire easy to understand. Some however, reported finding ambiguity in some wordings. For example, they queried what was meant by 'all students' (items 1, 2, 3, 10 and 13) and 'regular classroom' (items 1, 13, 14, and in the instruction of Part Two), stating that the understanding of these wordings would be subjective. They suggested that additional explanations in parentheses for these would be helpful. One item asked about whether the respondent knows anyone who has a disability, with clarification given in brackets, and one teacher interviewed found this particularly helpful and used it as an example as to how to improve clarity of other items in this way. One teacher also suggested that terminology such as 'SEND' and 'inclusion' should be clearly defined for the purposes of the questionnaire.

Information retrieval

Teachers reported that it was relatively easy to retrieve the information required to answer the items. Respondents with leadership and teaching responsibilities were able to answer the items, although some of their responses were influenced by their additional role rather than necessarily by their teaching role. A headteacher who held no teaching responsibilities suggested that their responses would be would probably answer the items differently if they responded as a teacher.

Judgement

Most teachers made comments about the options offered in the introductory items, suggesting that the questionnaire is amended to allow for the selection of more than one response options for some questions; for example, most schools in the local authority area are secondary schools with a sixth form yet respondents were not able to tick both answers. There were also queries about the number of options given for some of the numerical questions; for example, the question asking about the age range of respondents offered '40+' as an option, while other age-brackets were narrower, which *"...may have an effect on the answers given by someone who is in their early 40s and someone who is approaching retirement"*. Similarly, respondents suggested that there are several schools in the area of nearly 2000 students, which, they felt was quite different to the '1200+' option available. Another teacher commented that there did not seem to be sufficient options available in the question about the role(s) respondents have in their school, although they acknowledged these can vary widely from school to school.

Response

Most teachers found it straightforward to respond to all questions and would answer in the same way should they complete the questionnaire again.

There was some confusion however, with regard to response options such as 'none', 'some' and 'high' to the item asking respondent about how much exposure the respondent has had to students who have Special Educational Needs and Disabilities (SEND). Teachers reported that they were unsure what was meant by 'some' or 'high' and that interpretation of these words is subjective.

Further, one respondent was new to the school and commented that he found it easier to answer the questions about his own values than he did about his colleagues or the school in general. This issue was mitigated somewhat by the question at the end, '*How confident are you that your responses about your colleagues are accurate?*'. A question asking respondents how long they have been teaching in their current school might be useful to provide context for their answers, especially if the respondent sample is small and results could be influenced more by such answers. Respondents also suggested that a free-text box at the end for 'Other comments' may be helpful, or further questions about how confident they feel about all their answers or for any other comments they wished researchers to be aware about/know.

Finally, it was reported that there are many staff within secondary schools who work with students and are therefore involved in inclusion. Because the questionnaires focuses on teachers, it seems to exclude much of support staff who are crucial to include in order for one to understand inclusion in schools.

Layout and structure / Functionality

Teachers reported that they largely found the progress bar helpful, although one commented that it seemed to jump disproportionately. They also generally found the headings useful, framing the questions "*nicely*" and that they didn't influence responses.

Some of the sets of questions were rather long, it was said, and on a smaller screen it made it difficult to see the scale numbers at the top; this could be resolved by breaking down the sets of questions into smaller chunks for ease of use. Similarly, scrolling was also required for the 1-7 scales to ensure that respondents could see all the options, although this would not be possible to avoid on the survey platform used by the research team.

Timing

In total, 165 educators completed the questionnaires from all ten schools who consented to take part, although a number began the survey and dropped out before completion – the majority of which were on or after the first page.

The research team estimated that the teacher questionnaire would take about 20 minutes to complete, and most interviewees reported that this was accurate and that it was time well spent. Some raised the possibility that those who completed the survey already had an interest in inclusion, which is why they were keen to invest the time, and perhaps those who were less interested would not have done so in which case this may impact on the results.

Given teachers' time is limited, respondents suggested either: a) identifying the most important items that researchers wished teachers to respond to and instructing teachers to drop out after completing those if they wanted to but continue filling them in if they wanted to; and b) maybe creating a series of mini-surveys that would need less time to complete. Lastly, a respondent recommended that pop-up messages to show appreciation for respondents' time and to encourage them to continue could be added – they suggested that such messages were used in a survey they had recently completed and which they found to be motivating.

Bringing qualitative and quantitative data together

As it has already been indicated by the quantitative analysis, some of the items of the scale may have certain shortcomings, probably caused by the number of options used in the response scale: for a scale of six steps (1-6) some teachers would prefer to have the option for a 'neutral' or 'N/A' option, while others felt that not having this pushed them towards a 'truer' answer (respondents looked at options, wanting a neutral answer, but then had to decide between binary options and then to what extent)

Regarding the length of time working in the school: it is not clear how this may have affected answers in any way (see similar discussion for previous scale).

During interviews, some teachers suggested that there might be some ambiguity of wording. The qualitative data do not, however, provide any specific information on which items or terms in this scale may have been ambiguous.

It is important, however, to note that if any changes are attempted, the instrument would need to be re-piloted and re-evaluated for its psychometric characteristics.

Section 4 – Summary of Recommendations

A number of items (Items 7, 8, 11, 14, 15, 17, 18) may need to be removed from the analysis because they load on more than one factors or they have very low loadings (but also see pertinent discussion below). Removing the items would shorten the scale substantially, but we are still able to successfully perform an Exploratory Factor Analysis to reveal the structure of the instrument.

The analysis revealed two factors: 'Efficacy in inclusive instruction and collaboration' and 'Efficacy in managing behavior'. Essentially, the first factor of the TEIP-C scale combines two factors of the TEIP scale.

The TEIP-C scale yielded a different structure compared to the TEIP scale (two factors instead of three). It is not clear why this may have happened, and it would be interesting in a future research to contrast the responses of teachers on similar questions in TEIP and TEIP-C. This may also be the topic of some theoretical research.

Overall, there is a large number of items which loaded on more than one factors or had overall low loadings on factors. These items were removed from the analysis. The authors of the TEIP-C questionnaire might wish to review these items and decide on whether they should be part of any future administrations of the instrument.

Re-analysis of the data using a conceptual framework to group different items in categories in advance of the analysis might yield slightly different results to those in this report. We propose that such analysis is conducted before any changes are made to the questionnaire to ensure that items are not removed unnecessarily before other analysis methods are attempted to and the reliability of different items is tested in different ways.

The questionnaire could make it clear that respondents should only answer in their capacity as teachers for all questions as best as they can and should not fill in the questionnaire if they have no teaching responsibilities. Consideration could be given to other staff in the school, especially teaching assistants, as, in England, they are crucial to the everyday running of classrooms and of the schools.

Introductory questions could be revised to ensure that school information that can be found in publicly available databases are omitted from the questionnaire. This can reduce the items and thus participant burden and time filling in the questionnaire.

Overall, reducing items and creating a more user-friendly survey will be beneficial. For example, a mobile friendly or mobile first survey could be considered and could potentially improve response rates.

Bibliography

- Avramidis, E., & Norwich, B. (2002). Teachers' attitudes towards integration/inclusion: a review of the literature. *European journal of special needs education*, 17(2), 129-147.
- Baglin, J. (2014). Improving Your Exploratory Factor Analysis for Ordinal Data: A Demonstration Using FACTOR. *Practical Assessment, Research, and Evaluation*: 19(5), pp.NA.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309–319.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98–104
- De Boer, A., Pijl, S. J., & Minnaert, A. (2010). Attitudes of parents towards inclusive education: A review of the literature. *European Journal of Special Needs Education*, 25(2), 165-181.
- Dillman, Don A., Smyth, Jolene D., Christian, Leah Melani. (2014) *Internet, Phone, Mail and Mixed-Mode Surveys: The Tailored Design Method*, 4th edition. John Wiley: Hoboken, NJ.
- DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research, and Evaluation*, 14(1), 20.
- Nunnally, J. C. (1994). *Psychometric theory* 3E. Tata McGraw-hill education.
- R Core Team (2021). *R: A language and environment for statistical computing*. R. Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Paseka, A., & Schwab, S. (2020). Parents' attitudes towards inclusive education and their perceptions of inclusive teaching practices and resources. *European Journal of Special Needs Education*, 35(2), 254-272.
- Qvortrup, A., Qvortrup, L. (2018) Inclusion: Dimensions of inclusion in education, *International Journal of Inclusive Education*, 22:7, 803-817, DOI: 10.1080/13603116.2017.1412506.
- Revelle, W. (2021) *psych: Procedures for Personality and Psychological Research*, Northwestern University, Evanston, Illinois, USA, <https://CRAN.R-project.org/package=psych> Version = 2.1.9.
- Revelle, W., & Rocklin, T. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research*, 14(4), 403-414.
- Schueler, B. E., Capotosto, L., Bahena, S., McIntyre, J., & Gehlbach, H. (2014). Measuring parent perceptions of school climate. *Psychological assessment*, 26(1), 314.

- Schwab, S., Sharma, U., & Loreman, T. (2018). Are we included? Secondary students' perception of inclusion climate in their schools. *Teaching and Teacher Education*, 75, 31-39.
- Schwab, S., Sharma, U., & Hoffmann, L. (2022). How inclusive are the teaching practices of my German, Maths and English teachers?—psychometric properties of a newly developed scale to assess personalisation and differentiation in teaching practices. *International Journal of Inclusive Education*, 26(1), 61-76.
- Sharma, U. & Jacobs, K. (2016). Predicting in-service educators' intentions to teach in inclusive classrooms in India and Australia. *Teaching and Teacher Education* 55, 13-23.
- Sharma, U., Sokal, L., Wang, M., & Loreman, T. (2021). Measuring the use of inclusive practices among pre-service educators: A multi-national study. *Teaching and Teacher Education*, 107, 103506.
- Sharma, U., Loreman, T., & Forlin, C. (2012). Measuring teacher efficacy to implement inclusive practices. *Journal of research in special educational needs*, 12 (1), 12-21.
- Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment*, 80(1), 99-103.
- Sointu, E. T., Savolainen, H., Lappalainen, K., & Lambert, M. C. (2017). Longitudinal associations of student–teacher relationships and behavioural and emotional strengths on academic achievement. *Educational Psychology*, 37(4), 457-467.
- Tourangeau R. Cognitive science and survey methods: a cognitive perspective. In: Jabine T, Straf M, Tanur J, Tourangeau R, editors. *Cognitive Aspects of Survey Design: Building a Bridge between Disciplines*. Washington, DC: National Academy Press; 1984. pp. 73–100.
- Willis, 2009) G. Cognitive aspects of survey methodology. In: Lavrakas P, editor. *Encyclopedia of Survey Research Methods*. Vol 2. Thousand Oaks, CA: Sage Publications; 2009. pp. 103–106.
- Willis GB, Artino AR Jr. (2013) 'What Do Our Respondents Think We're Asking? Using Cognitive Interviewing to Improve Medical Education Surveys'. *J Grad Med Educ*. 2013 Sep;5(3):353-6. doi: 10.4300/JGME-D-13-00154.1
- UNESCO (1994) The Salamanca Statement on Principles, Policy and Practice in Special Needs.

Appendices

Appendix 1 – EFA for Attitudes to Inclusion Scale

Factor Analysis using method = minres

Loadings:

	MR1	MR2
25.1		0.752
25.3		0.763
25.4		0.811
25.5	0.813	
25.6	0.901	
25.7	0.819	
25.8	0.728	

	MR1	MR2
SS loadings	2.677	1.812
Proportion Var	0.382	0.259
Cumulative Var	0.382	0.641

Standardized loadings (pattern matrix) based upon correlation matrix

	MR1	MR2	h2	u2	com
25.1	-0.03	0.75	0.54	0.46	1
25.3	0.05	0.76	0.63	0.37	1
25.4	-0.01	0.81	0.64	0.36	1
25.5	0.81	0.06	0.73	0.27	1
25.6	0.90	-0.05	0.76	0.24	1
25.7	0.82	0.02	0.69	0.31	1
25.8	0.73	-0.02	0.51	0.49	1

	MR1	MR2
SS loadings	2.69	1.82
Proportion Var	0.38	0.26
Cumulative Var	0.38	0.65
Proportion Explained	0.60	0.40
Cumulative Proportion	0.60	1.00

With factor correlations of

	MR1	MR2
MR1	1.00	0.62
MR2	0.62	1.00

Mean item complexity = 1

Test of the hypothesis that 2 factors are sufficient.

The degrees of freedom for the null model are 21 and the objective function was 3.9 with Chi Square of 627.84

The degrees of freedom for the model are 8 and the objective function was 0.16

The root mean square of the residuals (RMSR) is 0.03

The df corrected root mean square of the residuals is 0.05

The harmonic number of observations is 164 with the empirical chi square 5.51 with prob < 0.7

The total number of observations was 165 with Likelihood Chi Square = 25.88 with prob < 0.0011

Tucker Lewis Index of factoring reliability = 0.922

RMSEA index = 0.116 and the 90 % confidence intervals are 0.068 0.168

BIC = -14.97

Fit based upon off diagonal values = 1

Measures of factor score adequacy

MR1 MR2

Correlation of (regression) scores with factors 0.95 0.92

Multiple R square of scores with factors 0.91 0.84

Minimum correlation of possible factor scores 0.81 0.68

Appendix 2 – EFA for Intention to Teach in Inclusive Classrooms (ITIC)

Factor Analysis using method = minres

Call: fa(r = d, nfactors = 2, n.obs = nrow(d), rotate = "oblimin",
fm = "minres", cor = "poly")

Standardized loadings (pattern matrix) based upon correlation matrix

	MR1	MR2	h2	u2	com
26.1	-0.15	0.65	0.33	0.67	1.1
26.2	0.84	0.02	0.73	0.27	1.0
26.3	0.88	-0.02	0.75	0.25	1.0
26.4	0.57	0.06	0.38	0.62	1.0
26.6	0.09	0.75	0.65	0.35	1.0
26.7	0.05	0.63	0.43	0.57	1.0

	MR1	MR2
SS loadings	1.86	1.40
Proportion Var	0.31	0.23
Cumulative Var	0.31	0.54
Proportion Explained	0.57	0.43
Cumulative Proportion	0.57	1.00

With factor correlations of

	MR1	MR2
MR1	1.00	0.58
MR2	0.58	1.00

Mean item complexity = 1

Test of the hypothesis that 2 factors are sufficient.

The degrees of freedom for the null model are 15 and the objective function was 2.09 with Chi Square of 337.16

The degrees of freedom for the model are 4 and the objective function was 0.04

The root mean square of the residuals (RMSR) is 0.02

The df corrected root mean square of the residuals is 0.04

The harmonic number of observations is 164 with the empirical chi square 2.51 with prob < 0.64

The total number of observations was 165 with Likelihood Chi Square = 6.74 with prob < 0.15

Tucker Lewis Index of factoring reliability = 0.968

RMSEA index = 0.064 and the 90 % confidence intervals are 0 0.147

BIC = -13.69

Fit based upon off diagonal values = 1

Measures of factor score adequacy

	MR1	MR2
Correlation of (regression) scores with factors	0.93	0.88
Multiple R square of scores with factors	0.87	0.77
Minimum correlation of possible factor scores	0.74	0.54

Appendix 3 – EFA for Teacher Efficacy for Inclusive Practice (TEIP)

Factor Analysis using method = minres

Standardized loadings (pattern matrix) based upon correlation matrix

	MR1	MR2	MR3	h2	u2	com
27.2.	0.07	-0.08	0.65	0.43	0.566	1.1
27.3.	-0.13	0.33	0.66	0.58	0.419	1.6
27.4.	0.22	-0.06	0.72	0.70	0.301	1.2
27.5.	0.05	0.03	0.67	0.50	0.500	1.0
27.7.	0.98	-0.04	-0.04	0.88	0.123	1.0
27.8.	0.91	0.07	0.00	0.90	0.096	1.0
27.9.	0.83	0.13	-0.02	0.80	0.197	1.0
27.10	0.91	0.02	0.05	0.90	0.102	1.0
27.12	0.71	-0.06	0.24	0.68	0.317	1.2
27.16	0.18	0.76	-0.09	0.70	0.301	1.1
27.17	0.03	0.85	0.09	0.83	0.174	1.0
27.18	-0.02	0.94	0.01	0.87	0.131	1.0

	MR1	MR2	MR3
SS loadings	4.18	2.49	2.11
Proportion Var	0.35	0.21	0.18
Cumulative Var	0.35	0.56	0.73
Proportion Explained	0.48	0.28	0.24
Cumulative Proportion	0.48	0.76	1.00

With factor correlations of

	MR1	MR2	MR3
MR1	1.00	0.53	0.55
MR2	0.53	1.00	0.37
MR3	0.55	0.37	1.00

Mean item complexity = 1.1

Test of the hypothesis that 3 factors are sufficient.

The degrees of freedom for the null model are 66 and the objective function was 11.15 with Chi Square of 1774.69

The degrees of freedom for the model are 33 and the objective function was 0.61

The root mean square of the residuals (RMSR) is 0.03

The df corrected root mean square of the residuals is 0.04

The harmonic number of observations is 164 with the empirical chi square 13.67 with prob < 1

The total number of observations was 165 with Likelihood Chi Square = 95.13 with prob < 6.2e-08

Tucker Lewis Index of factoring reliability = 0.926

RMSEA index = 0.107 and the 90 % confidence intervals are 0.082 0.133

BIC = -73.37

Fit based upon off diagonal values = 1

Measures of factor score adequacy

	MR1	MR2	MR3
Correlation of (regression) scores with factors	0.98	0.97	0.91
Multiple R square of scores with factors	0.97	0.93	0.84
Minimum correlation of possible factor scores	0.94	0.87	0.67

Appendix 4 – EFA for Teacher Efficacy for Inclusive Practice - Collective (TEIP-C)

Factor Analysis using method = minres

Call: fa(r = d, nfactors = 2, n.obs = nrow(d), rotate = "oblimin",
fm = "minres", cor = "poly")

Standardized loadings (pattern matrix) based upon correlation matrix

	MR1	MR2	h2	u2	com
28.1.	0.85	-0.03	0.68	0.32	1.0
28.2.	0.75	0.11	0.68	0.32	1.0
28.3.	1.00	-0.13	0.84	0.16	1.0
28.4.	0.76	0.18	0.79	0.21	1.1
28.5.	0.62	0.22	0.61	0.39	1.3
28.6.	0.54	0.19	0.46	0.54	1.3
28.9.	0.13	0.67	0.58	0.42	1.1
28.10	0.07	0.79	0.70	0.30	1.0
28.12	-0.07	0.89	0.72	0.28	1.0
28.13	0.12	0.73	0.67	0.33	1.1
28.16	-0.04	0.79	0.59	0.41	1.0

	MR1	MR2
SS loadings	3.85	3.47
Proportion Var	0.35	0.32
Cumulative Var	0.35	0.67
Proportion Explained	0.53	0.47
Cumulative Proportion	0.53	1.00

With factor correlations of

	MR1	MR2
MR1	1.00	0.64
MR2	0.64	1.00

Mean item complexity = 1.1

Test of the hypothesis that 2 factors are sufficient.

The degrees of freedom for the null model are 55 and the objective function was 9.06 with Chi Square of 1445.25

The degrees of freedom for the model are 34 and the objective function was 1.11

The root mean square of the residuals (RMSR) is 0.05

The df corrected root mean square of the residuals is 0.06

The harmonic number of observations is 163 with the empirical chi square 44.91 with prob < 0.1

The total number of observations was 165 with Likelihood Chi Square = 174.88 with prob < 7.2e-21

Tucker Lewis Index of factoring reliability = 0.835

RMSEA index = 0.158 and the 90 % confidence intervals are 0.136 0.183

BIC = 1.28

Fit based upon off diagonal values = 0.99

Measures of factor score adequacy

	MR1	MR2
--	-----	-----

Correlation of (regression) scores with factors	0.97	0.96
Multiple R square of scores with factors	0.94	0.91
Minimum correlation of possible factor scores	0.88	0.83