‘Are We Included?’ A validation of the students’ Inclusion Climate Scale (ICS)

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 students who enthusiastically filled in the Inclusion Climate Scale (ICS) questionnaire and provided invaluable insights during the interview process.

Contents

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.

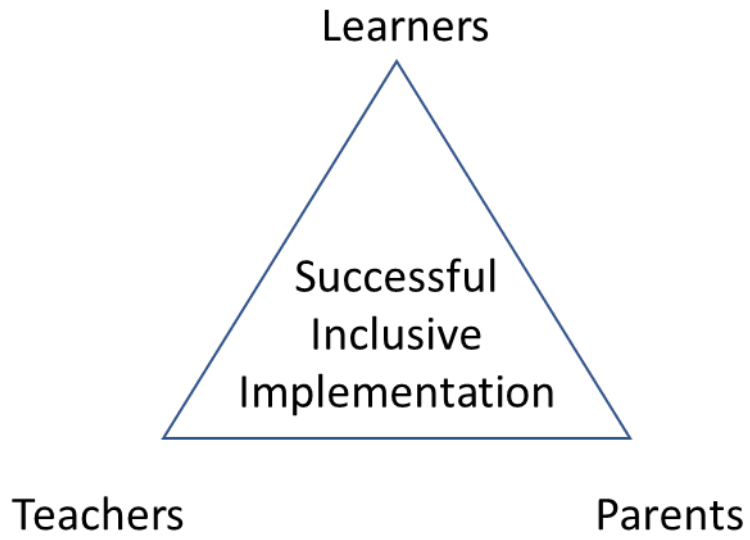


Figure 1. Tree important agents of inclusion education: parents, students and teachers.

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 learner attitudes, the research uses scales which have been developed and tested relatively recently (for example, see Schwab et al, 2018).

Section 2 – Data and Methods

Aims and Objectives

This study aimed to:

a) validate the Inclusion Climate Scale (ICS) in the context of secondary schools in Plymouth, England

b) compare its findings with those of Schwab, Sharma, & Loreman (2018), who validated the instrument in 18 secondary schools in North Rhine-Westphalia, a federal state in Germany, and found that it demonstrated satisfactory reliability and validity.

The study is important as it assesses the validity and reliability of the scale 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 Schwab, Sharma, & Loreman (2018), the research community would be encouraged to use the scale across the UK or across countries. Similar findings from two different European countries would suggest that the scale produces valid and reliable results across educational systems and cultures.

Data and Methods

The questionnaire

We used the same instrument which was originally developed and used by Schwab, Sharma, & Loreman (2018). The scale consists of 28 items (see Appendix 1) probing the attitudes of students regarding the inclusion climate at their school. The instrument uses a four-step scale: 1. Not at all true, 2. Slightly true, 3. Very true, 4. Completely true.

Items 9, 21, 24 and 25 are negatively worded; see, for example, Item 21 “My teachers are not very keen with teaching students who are shy and withdrawn." Items with negative wording need to be reversed back to a positive statement (e.g. “My teachers are very keen …” so as not to have negative correlations with the other items of the scale. Sometimes, negatively worded items may also be difficult for young individuals to comprehend and respond to. Specifically, Schwab, Sharma, & Loreman (2018) suggested that “… students, even if they are in secondary schools, struggled with negatively formulated items. All negatively worded items were eliminated from our scale.” (p. 37). At this stage it is not clear if the same phenomenon will be observed in our data.

The questionnaire was constructed online using JISC. A link to the online questionnaire was distributed to schools, which disseminated these to their Year 7 and Year 10 students. Online questionnaires were initially set to be open for two weeks. During the first two weeks, take-up was relatively slow, so the online surveys were kept open for a further two weeks. To encourage participation weekly reminders were built into the process, using Dillman’s Tailored Design Method (2014), which proved effective.

Schools decided on how best to give students access to the online questionnaire via computer; some did so in tutor groups using the ICT suite, others facilitated the completion en masse with whole year groups at a time, while others offered it to single students at a time that suited them. A school staff was present to check if students needed any help with the questionnaire, as well as observe if any of the

respondents was experiencing distress and act accordingly. All responses were anonymous.

As regards the analysis, Schwab, Sharma, & Loreman (2018) used exploratory factor analysis (EFA) to reveal two factors in the data: (a) Factor I: Teacher Support and Care; and (b) Factor II: Emotional Experience. Factor I ('Teacher support and care') consisted of the following items: 5, 10, 14, 16, 17, 19, 23, 26, 27. Factor 2 ('Emotional Experience') consisted of the following items: 1-4, 6.

We used EFA to reveal the underlying factors in the data. We did not remove the negatively worded items before the analysis. Due to the ordinal nature of the scale "Not at all true" to "Completely true", we used polychoric correlations to compute the correlation matrix for the EFA.

Schwab, Sharma, & Loreman (2018) investigated the discriminating power of items using the up/low 27% method. We chose to use standard psychometric techniques (e.g. item-total correlations) which are equally informative and will probably be more familiar to the readers of the report.

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 using a sub-sample of 833 students who gave responses to all 28 items of the scale (74 students yielded one or more missing responses). 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). We also used oblimin rotation allowing the factors to correlate between them.

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 three 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; of the 907 questionnaire

respondents, 453 provided their email address. These students were contacted by the research team via their school email address to confirm participation and, of those who responded first, we posed two selection criteria: that the sample covered Year 7 and Year 10 students and that these students attended different schools.

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

Overall, 907 students from 9 schools located in the city of Plymouth filled in the questionnaire and three students were interviewed: two from Year 7 and one from Year 10. These students attended different schools.

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 students and their parents by the school. A two-week window was offered to parents to opt out their children from the study. Students were asked to provide their consent before completing the questionnaire and before the interviews started. Withdrawal from the study as a whole was possible until July 2022, when the analysis phase begun.

Given that some children may have been identified as vulnerable by the school, as well as the potential distress that some children might feel due to the nature of the questions asked in the questionnaire and interviews, Headteachers and Inclusion Leads were asked to use their professional judgement in ensuring children who took part in the questionnaire were not in risk of experiencing harm. Headteachers and Inclusion Leads were also asked to make their pastoral teams aware of when students would be completing the questionnaires and taking part in interviews. This

was to ensure that additional pastoral care would be provided to any students who became distressed as a result of taking part in the study. A school staff was also asked to be present whilst students were completing the questionnaire to ensure that all goes smoothly but also to observe if any of the respondents was experiencing distress and act accordingly.

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

Section 3 - Results

Out of the 907 completed questionnaires, circa 52% were Year 7 students and the rest were Year 10 students (there were 3 Year 11 students). Approximately half of the students were boys. The students' age ranged between 11-15 (there were 2 students aged 17), with a mean age of 13.2 (sd=1.58).

Descriptive statistics

Initially, we inspected the frequency distributions for all 28 items of the scale (see Appendix 2). Some of the items seem to demonstrate less variance than others. For example, Item 12 ("I have at least one friend in my school who cares about me") has limited variance as almost all of the responses given were '4 Completely true'. Other items demonstrate much higher variance and respondents utilized the whole range of the scale. For example, item 11 ("Most of my classmates like me") received responses across the whole range of the scale 1-4. Items with very little variance often do not correlate satisfactorily with other items and as a result may not be included in common factors during the EFA.

Item inter-correlations

First, we reversed items 9, 21, 24 and 25 to avoid artificially inducing negative correlations between the items. Figure 1 presents the correlogram of the polychoric correlations between the 28 items of the ICS. The figure suggests that certain items have particularly low correlations with the rest of the items. For example, items 11 ("Most of my classmates like me.") and 12 ("I have at least one friend on my school who cares about me.") have particularly low correlations with all other items but item 18 ("My classmates invite me to social events (e.g. birthday parties)."). Schwab, Sharma, & Loreman (2018) had removed those items because they had low correlations with the 'total score' (they referred to the item-total correlation) but we see no obvious reason to delete them because they have a relatively high correlation between them and it is likely that these three items may form a theoretically meaningful common factor (albeit with near-zero correlation with any other factors that may emerge).

Moreover, item 21 ("My teachers are not very keen on teaching students who are shy and withdrawn."), item 24 ("I have been bullied by other students at this school at least once") and item 25 ("Teachers are not interested in teaching students who frequently misbehave in class.") have very low correlations with all other items, as well as between them. These are three out of the four items which were negatively worded and which were originally removed from further analysis by Schwab, Sharma, & Loreman (2018). Item 9, the fourth negatively worded item also had small correlations with the rest of the items. Following the example of Schwab, Sharma, & Loreman (2018) we decided to remove these items (9, 21, 24, 25) from further analysis.

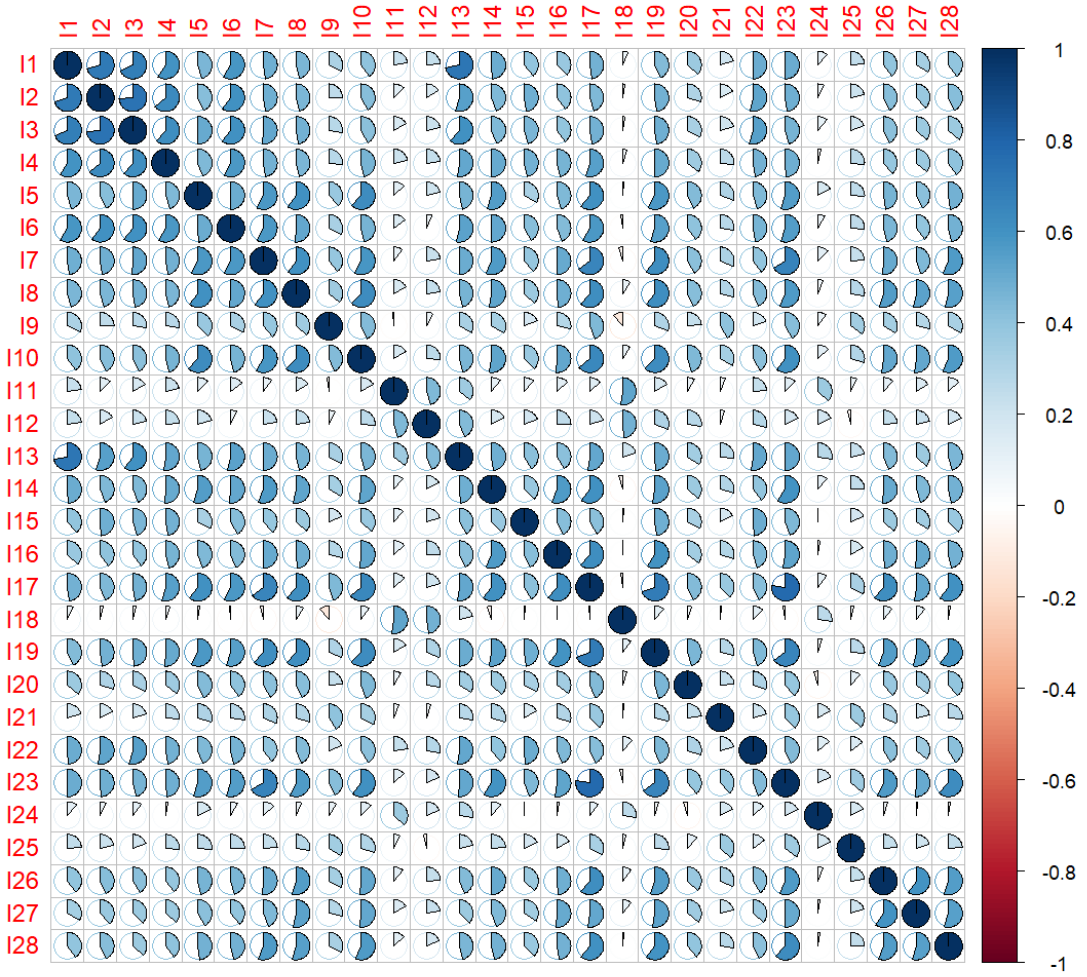


Figure 2. Correlogram of polychoric correlations between the 28 items of the scale. Larger proportion of shaded pie-charts (and darker shade) represent larger correlations.

The factor structure

A Parallel Analysis on the remaining 24 items (after removing the four negatively worded items), suggested the extraction of between 3 and 4 factors. VSS complexity 2 achieved a maximum of 0.92 with 2 factors whereas the Velicer MAP achieved a minimum of 0.01 with 3 factors. A scree plot suggested the extraction of 3 factors. We decided to run a preliminary EFA with 3 factors. The results are shown in Appendix 3. We investigated the matrix with factor loadings; some of the items (Items 6, 13, 15 and 22) have substantial loading on more than one factor so we decided to remove from the analysis and re-run the EFA. For the EFA we used oblimin rotation because we have every reason to believe that the main factors would be correlated.

The final EFA analysis yielded three factors, two of which resemble closely the factors found by Schwab, Sharma, & Loreman (2018) – see Appendix 4.

The first factor consists of 13 items: 5, 7, 8, 10, 14, 16, 17, 19, 20, 23, 26, 27, 28. This factor includes all the items suggested by Schwab, Sharma, & Loreman (2018),

with the addition of items 7, 8, 20, 28 and we agree that the name ‘Teacher support and care’ and is an appropriate name for this set of items.

The second factor consists of five items: 1, 2, 3, 4. These are the same items as the ones suggested by Schwab, Sharma, & Loreman (2018) with the deletion of item 6 (it was removed in a previous stage of the analysis). Again, we agree that an appropriate name for this set of items would be ‘Emotional Experience’.

The third factor consists of only three items: 11, 12, 18. The items are copied here for ease of reference:

- Item 11: Most of my classmates like me.
- Item 12: I have at least one friend on my school who cares about me.
- Item 18: My classmates invite me to social events (e.g. birthday parties).

The wording of the names suggests that an appropriate name for this set of items might be ‘Peer acceptance and support’.

The third factor, as predicted from the investigation of the correlation matrix, represents the three items which had been removed by Schwab, Sharma, & Loreman (2018) before the analysis. If Schwab et al hadn’t removed those items early on, it is very likely that they would have recovered this factor as well. So, overall, our findings corroborate past findings closely. It is very interesting to note that these are the only items (out of the 28 items of the scale) which refer to the classmates of the students

Interestingly, the factors ‘Emotional Experience’ and ‘Teacher support and care’ are highly correlated (r=0.68), whereas the correlation of the ‘Peer acceptance and support’ and the other two factors is minimal (less than 0.2). The fact that we have recovered a third factor (the effect of students) which is not related to the other two may suggest that the students indeed differentiate between teachers and students as possible determinants of the schooling experience.

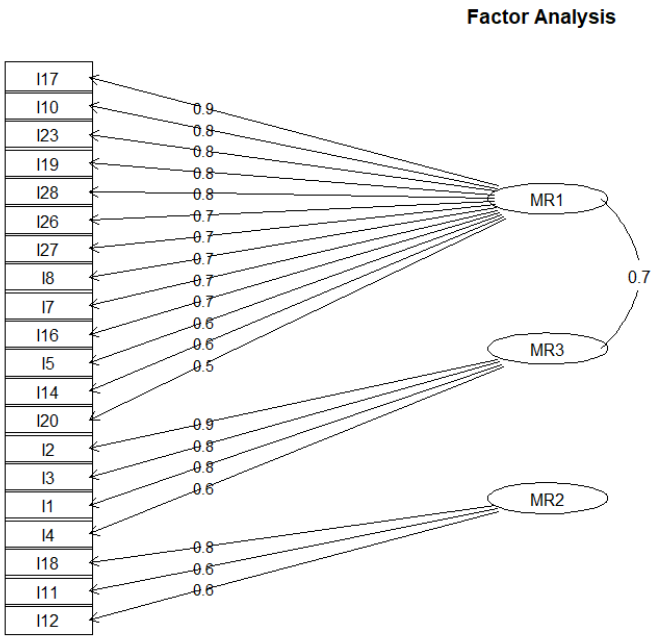


Figure 3. The factor structure of the scale.

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.

Important note: replicating the analysis separately for the Year 10 or Year 7 students does not produce a substantially different factor structure. By pursuing the same factor structure for the whole sample, we facilitate comparisons across year groups. Future research, however, depending on the aims of the study, could pursue a separate factor analysis for the two groups.

Reliability and Discriminating power of items

For each of the three factors, we computed Cronbach's alpha. The Cronbach's alpha for the first factor ('Teacher support and care'), based on polychoric correlation matrix, was 0.94. The corrected item-total correlation for the items of the factor ranged from 0.54 to 0.85 which is satisfactory for all intents and purposes, suggesting that all items contribute significantly to the measurement exercise. The average item inter-correlation is 0.54, 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 ('Emotional Experience'), based on polychoric correlation matrix, was 0.89. The corrected item-total correlation for the items of the factor ranged from 0.79 to 0.85 which is satisfactory for all intents and purposes. The average inter-item correlation was very high, at 0.66, which suggests that the items tap on the same construct, as intended.

The Cronbach's alpha for the third factor ('Peer acceptance and support'), based on polychoric correlation matrix, was 0.74. This is a very satisfactory value for Cronbach's alpha, for a scale 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). The corrected item-total correlation for the items of the factor ranged from 0.62 to 0.65 and 0.69. With an average inter-item correlation of 0.48, this scale is within the recommendations of Clark & Watson (1995). Overall, the three scales seem to have very satisfactory reliability indices and high inter-item correlations.

Findings from the interviews

Comprehension

Overall, interviewees reported to have understood most items with relative ease. There were some items however, which seem to have inhibited students' reading comprehension. In particular, some of the vocabulary used in these items, including the words 'majority' (items 3, 9 and 16), 'subject' (items 10, 15 and 22), 'impartial' (items 14) and 'co-operation' (items 26 and 27), was perceived to be relatively impenetrable. Items 16 and 36 were also seen unsupportive of students' understanding because they used long sentence and/or had syntax that was hard to follow.

Further, interviewees found items 22, 26, 27, 35 and 36 in need of further defining and/or clarification. In more detail, item 22 asks students to rate their satisfaction with their school 'achievement' in most of their subjects. One interviewee questioned what type of achievement they had to respond about - academic, personal or any other? In a similar vein, one interviewee pointed to the need to further define words such as

'satisfaction' (items 35 and 36), which might mean different things to different people. For items 26 and 27 ('Teachers encourage co-operation among students,' and 'Teachers in our school co-operate effectively with each other'), students struggled to 'translate' their theoretical understanding of the meaning of the word 'co-operation' into how it might be 'enacted'/ applied in practice and within a particular context. For example, although one of the three interviewees provided the correct meaning of the word co-operation, they were unsure how co-operation between teachers might be exemplified. This led them to start 'second guessing' themselves and ponder about how they should answer the question. The same student suggested that including concrete examples could scaffold students' thinking/support their understanding on how to answer such items. Interestingly, apart from item 35, all the items identified as needing further defining/clarification were also highlighted as having a word which students could not 'access' or having a hard-to-follow syntax.

Question	Challenge	How many students
3, 9 & 16	• *'majority'* - meaning of the word unknown/challenging. suggestion to replace with 'most'	• 1 student (Year 7)
10, 15 & 22	• *'subject'* - meaning of the word unknown/challenging	• 1 student (Year 7)
13	• The question asks students to rate how happy they are in the schools. Suggestion to add a question that asks students whether they think if most of their classmates are happy too	• 1 student (Year 10)
14	• *'impartial'* - meaning of the word unknown/challenging	• 2 students (Year 7 and 10)
16	• Long sentence/syntax making it challenging to understand	• 1 student (Year 10)
26	• *'co-operation'* - meaning of the word unknown/challenging • Clarity about how co-operation between students might look like. Suggestion to add examples	• 2 students (Year 7 and Year 10) • 1 student (Year 10)
27	• *'co-operation'* - meaning of the word unknown/challenging • Clarity about how co-operation between teachers might look like. Suggestion to add examples	• 2 students (Year 7 and 10) • 1 student (Year 10)
32	• Challenging for students to identify what type of school they were attending (e.g. selective, comprehensive, etc.)	• 1 student (Year 10)
35	• Clarity /defining what satisfaction means in the context of the questionnaire	• 1 student (Year 10)
36	• Clarity /defining what satisfaction means in the context of the questionnaire • Syntax and/or meaning unclear	• 1 student (Year 10)

Table 1. Summary of learner feedback on items perceived to need further defining / clarification.

Information retrieval

The questionnaire does not set a timeframe for information retrieval. Interviewees reported answering items based on notable school events of the past few weeks or months. This meant, they suggested, that their responses were a snapshot in time and could change over time.

Judgment

As expected, how easy or difficult to derive answers for the items was influenced by students' comprehension. Interviewees also reported lacking knowledge of how schools are categorized in England, prohibiting them from easily identifying whether the schools they were attending was, for example, a mainstream secondary school, or special school, or a special class in a mainstream school.

Response

There was agreement that 'time' and more recent experiences in the school influence how respondents answered the items. Concerns were also raised about the four-point scale used in the questionnaire missing a mid-point, resulting to some less 'truthful' answers, it was reported. There were also questions raised about the effectiveness of a fou- point scale in capturing the extent of change over time. One interviewee also felt that longer, i.e., a 10-point scale, would support students to gauge "*easier where I would lie; in the middle or at the high end or at the lower end or 1/3 of the way to the middle"*. When the same student was asked to use a 10 point scale to answer some of the items randomly, they rated items 3 and 6, 7 and 9 respectively from their original 3.

Timing

Two of the interviewees reported completing the questionnaire within 15 minutes (year 7s) whilst one needed 25 minutes (Year 10). There was consensus that the time spent was worthwhile and interviewees were happy to re-invest that time to. complete the questionnaire again in the future.

Layout and structure / Functionality

Two interviewees suggested that the layout and structure of the online questionnaire were appropriate and relatively easy to follow, whilst one disagreed, suggesting that answering each item individually was time consuming. They suggested laying out items in a list. It was also suggested that the questionnaire's introductory text could be replaced with a video using students to present the details of the study, alongside their own experience of filling the questionnaire in to further enhance engagement.

There was consensus that functionality was relatively unproblematic. The progress bar included in the questionnaire appears to have motivated students to continue filling in the questionnaire, but it was not necessarily accurate in showing the progress they were making, according to one interviewee.

Bringing qualitative and quantitative data together

As it has already been indicated by the qualitative analysis, some of the items of the scale may have certain shortcomings. For example, interview data suggested that the wording of item 22 ("I am satisfied with my school achievement in most of my subjects") might not be age-appropriate as some students might not understand the word 'subject' and/or other were unclear about type of 'school achievement' being asked. The quantitative analysis has also shown that this particular item should be removed from further analysis as it loaded on two different factors; although this

might or might not be related to the wording problems, it is a fact that item 22 could be removed from the scale without any significant loss of information.

In some other cases, during interviews, students suggested that the wording, length of the sentence or its syntax might also need to be more age-appropriate, specifically in relation to these items: 3, 9, 16 ('majority'), 10, 15 and 22 ('subject'), 14 ('impartial'), 26 and 27 ('co-operation'), 16 and the two overall 'satisfaction' items (36 and 37), which are not formally part of the ICS. The quantitative analysis has not revealed any significant flows for these items.

Other qualitative feedback suggested that using a scale with more 'steps' (e.g. 1 – 10) might work better than just four options. The quantitative data has not revealed any significant problems; on the contrary the three sub-scales (i.e., factors) identified have desirable psychometric characteristics.

Other suggestions emerging from the qualitative analysis are to have additional information such as definitions, clarifications or examples to explain what is meant by some words and sentences. It is important, however, to note that if such changes are attempted, the instrument would need to be re-piloted and re-evaluated for its psychometric characteristics.

Section 4 – Summary of Recommendations

Some of the items of the ICS are not needed and might be removed from future administrations of the instrument (Items 6, 9, 13, 15, 21, 22, 24, 25). Removing these items would shorten the scale which would save a significant amount of time for the respondents (i.e., the students), the analysis time for the researchers and practitioners, and would also yield a more compact but solid instrument.

We also propose that some further cultural adaptations take place, including reviewing: a) the options related to the demographic information asked, particularly the options offered under gender which were identified as restrictive in the qualitative interviews; b) the options in terms of type of school. This item it could be removed as, if students select or type the name of their school, researchers could match the students' answer with publicly available school information.

For purposes of further analysis, we recommend to aggregate three 'scores' for each student, based on the three factors revealed: 'Emotional Experience', 'Teacher support and care' and 'Peer acceptance and support'. These three scores could be used to investigate differences between groups with different demographic characteristics or between schools etc.

Replicating the analysis separately for the Year 10 or Year 7 students does not produce a substantially different factor structure. By pursuing the same factor structure for the whole sample, we facilitate comparisons across year groups. In a future research, however, depending on the aims of the study, it may be desirable to pursue a separate factor analysis for the two groups

The scale has very satisfactory psychometric characteristics and our results have largely corroborated the finding of Schwab, Sharma, & Loreman (2018), with the addition of the third factor 'Peer acceptance and support' which is probably our most important finding. It would be interesting to re-analyze the data by Schwab, Sharma, & Loreman (2018) to investigate whether this factor can be recovered. There were however some concerns raised by the use of a four and fine point scales in terms of effectively capturing the extend of change over time which we recommend are considered.

If changes are attempted on the wording of items and questionnaire layout, the instrument would need to be re-piloted and re-evaluated for its psychometric characteristics.

We also propose that the item asking students to identify which school they attend is removed; students were asked to select their school out of a list. If the school names are known to the researchers wither by students clicking an option or by typing it in, researchers could match the students' answer with publicly available UK school information.

Lastly, 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. URLhttps://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.

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.

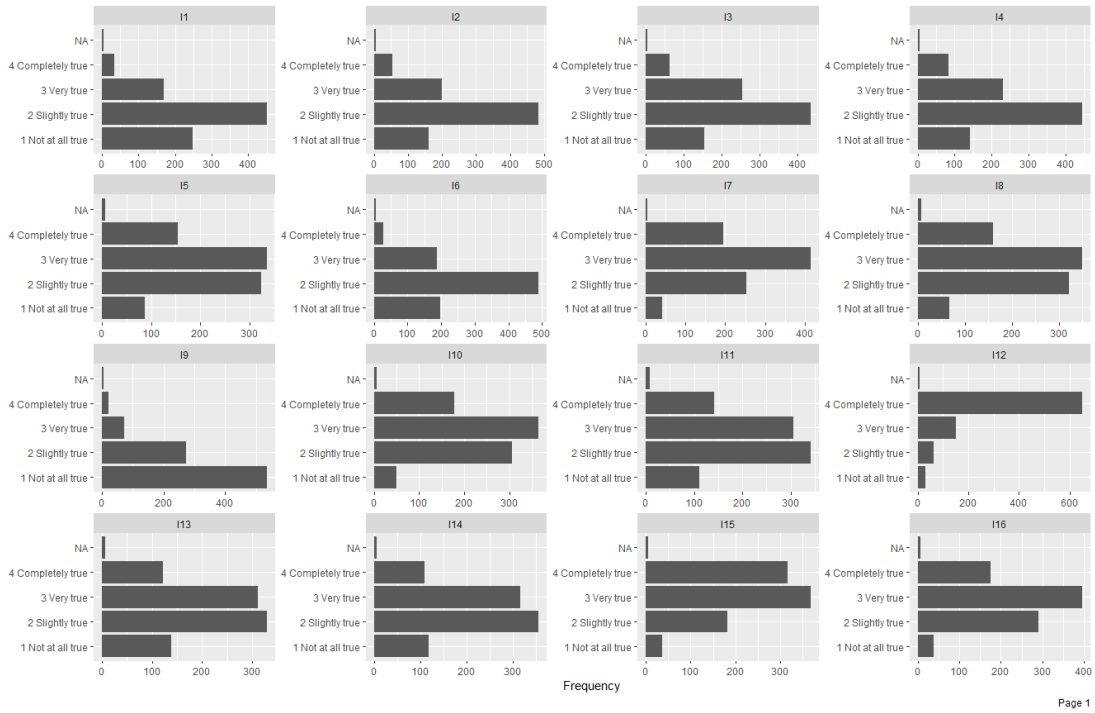
Appendices

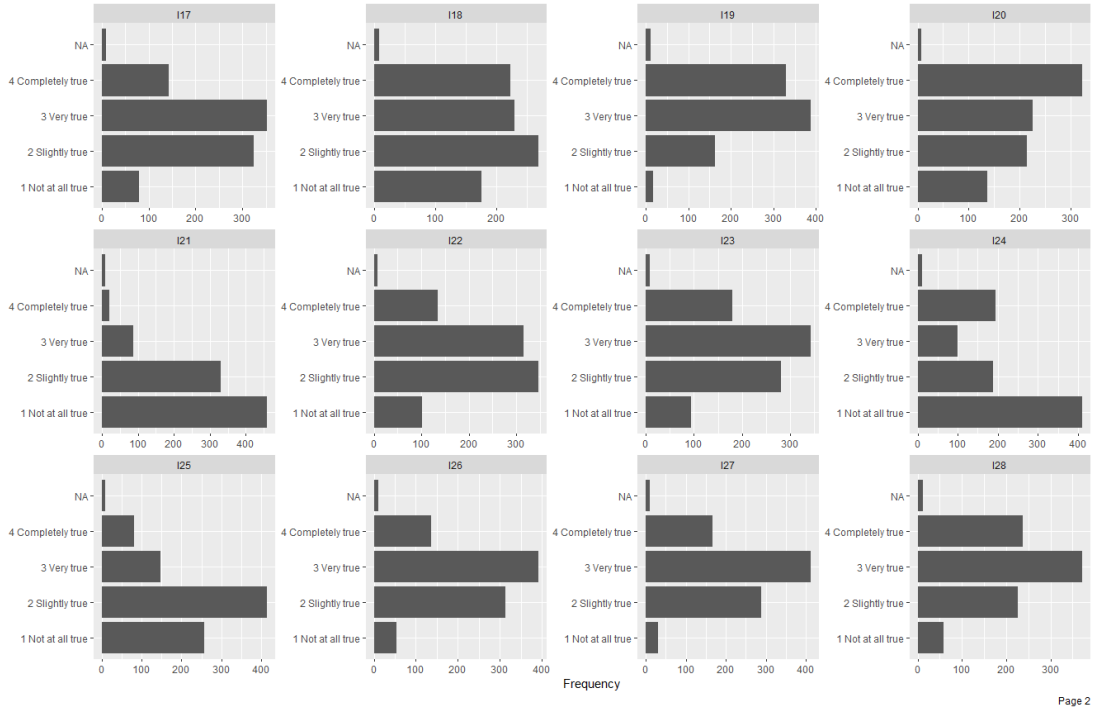
Appendix 1 – The instruments

		1 Not at all true	2 Slightly true	3 Very true	4 Completely true
1	I enjoy coming to school every day.				
2	I enjoy attending most of my classes.				
3	I like the majority of lessons in my school.				
4	I look forward to participating in classroom activities that my teachers have planned for us.				
5	I receive enough help from teachers if I struggle to do school work.				
6	Most teachers in my school make learning fun.				
7	Teachers and other staff in this school are friendly to me.				
8	My teachers give me supportive feedback when I do well at school.				
9	The majority of my teachers are not interested in teaching students who struggle with their learning.				
10	My teachers make sure that students, who face difficulty in learning a subject, receive enough support and guidance.				
11	Most of my classmates like me.				
12	I have at least one friend on my school who cares about me.				
13	I am happy at this school.				

14	My teachers are impartial and apply school rules in a fair way when somebody misbehaves in the class.				
15	I try to do my best in all subjects.				
16	My teachers make sure that all students actively participate in the majority of school and classroom activities.				
17	My teachers are very caring about all students.				
18	My classmates invite me to social events (e.g. birthday parties).				
19	My teachers want me to work as good as possible and to do well.				
20	I have at least one teacher in my school who I can contact if I am facing any difficulties.				
21	My teachers are not very keen on teaching students who are shy and withdrawn.				
22	I am satisfied with my school achievement in most of my subjects.				
23	Teachers treat all students with respect at my school.				
24	I have been bullied by other students at this school at least once.				
25	Teachers are not interested in teaching students who frequently misbehave in class.				
26	Teachers encourage co-operation among students.				
27	Teachers in our school co-operate effectively with each other.				
28	Teachers interact respectfully with all students' parents.				

Appendix 2– Frequency distributions per item





Note: 'NA' represents missing responses

Appendix 3 – Preliminary Exploratory Factor Analysis

```
Factor Analysis using method =  minres
Call: fa(r = dd, nfactors = 3, n.obs = nrow(dd), rotate = "oblimin",
    scores = "regression", fm = "minres", cor = "poly")


Loadings:
    MR1    MR3    MR2
I1          0.852
I2          0.864
I3          0.860
I4   0.172  0.610
I5   0.638  0.114
I6   0.303  0.536 -0.114
I7   0.681  0.136
I8   0.715
I10  0.813
I11         0.142  0.648
I12  0.165         0.649
I13  0.143  0.588  0.275
I14  0.608  0.163
I15  0.261  0.370
I16  0.660
I17  0.871
I18              0.771
I19  0.792
I20  0.494
I22  0.195  0.473  0.140
I23  0.786
I26  0.723
I27  0.720
I28  0.770


Factor Analysis using method =  minres
Call: fa(r = dd, nfactors = 3, n.obs = nrow(dd), rotate = "oblimin",
    scores = "regression", fm = "minres", cor = "poly")
Standardized loadings (pattern matrix) based upon correlation matrix
      MR1   MR3   MR2   h2   u2 com
I1  -0.04  0.85  0.05 0.69 0.31 1.0
I2  -0.02  0.86 -0.06 0.70 0.30 1.0
I3  -0.03  0.86 -0.01 0.71 0.29 1.0
I4   0.17  0.61  0.03 0.56 0.44 1.2
I5   0.64  0.11 -0.02 0.52 0.48 1.1
I6   0.30  0.54 -0.11 0.59 0.41 1.7
I7   0.68  0.14 -0.08 0.59 0.41 1.1
I8   0.72  0.05  0.05 0.57 0.43 1.0
I10  0.81 -0.07  0.07 0.61 0.39 1.0
I11 -0.04  0.14  0.65 0.46 0.54 1.1
I12  0.17  0.02  0.65 0.50 0.50 1.1
I13  0.14  0.59  0.28 0.63 0.37 1.6
I14  0.61  0.16 -0.08 0.52 0.48 1.2
```

```
I15  0.26  0.37  0.01 0.34 0.66 1.8
I16  0.66  0.04  0.01 0.48 0.52 1.0
I17  0.87 -0.01 -0.06 0.73 0.27 1.0
I18 -0.06 -0.05  0.77 0.57 0.43 1.0
I19  0.79  0.01  0.07 0.66 0.34 1.0
I20  0.49  0.05  0.06 0.30 0.70 1.0
I22  0.20  0.47  0.14 0.44 0.56 1.5
I23  0.79  0.06 -0.08 0.67 0.33 1.0
I26  0.72 -0.03  0.05 0.51 0.49 1.0
I27  0.72 -0.10  0.09 0.46 0.54 1.1
I28  0.77 -0.05 -0.01 0.54 0.46 1.0

                       MR1  MR3  MR2
SS loadings           7.50 4.19 1.66
Proportion Var        0.31 0.17 0.07
Cumulative Var        0.31 0.49 0.56
Proportion Explained  0.56 0.31 0.12
Cumulative Proportion 0.56 0.88 1.00

 With factor correlations of
     MR1  MR3  MR2
MR1 1.00 0.70 0.18
MR3 0.70 1.00 0.18
MR2 0.18 0.18 1.00

Mean item complexity =  1.2
Test of the hypothesis that 3 factors are sufficient.

The degrees of freedom for the null model are  276  and the objective function was
15.18 with Chi Square of  13622.65
The degrees of freedom for the model are 207  and the objective function was  1.21

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

The harmonic number of observations is  896 with the empirical chi square  422.74
with prob <  6.5e-17
The total number of observations was  907  with Likelihood Chi Square =  1083.58
with prob <  1.1e-118

Tucker Lewis Index of factoring reliability =  0.912
RMSEA index =  0.068  and the 90 % confidence intervals are  0.064 0.072
BIC =  -326.12
Fit based upon off diagonal values = 1
Measures of factor score adequacy
                                                MR1  MR3  MR2
Correlation of (regression) scores with factors   0.97 0.96 0.88
Multiple R square of scores with factors          0.95 0.92 0.78
Minimum correlation of possible factor scores     0.90 0.84 0.55
```

Appendix4 – Final Exploratory Factor Analysis with 3 factors

```
Factor Analysis using method =  minres
Call: fa(r = dd, nfactors = 3, n.obs = nrow(dd), rotate = "oblimin",
    scores = "regression", fm = "minres", cor = "poly")


Loadings:
    MR1    MR3    MR2
I1          0.800
I2          0.871
I3          0.844
I4   0.185  0.601
I5   0.641  0.117
I7   0.684  0.132
I8   0.710
I10  0.802
I11         0.124  0.644
I12  0.185         0.604
I14  0.615  0.156
I16  0.669
I17  0.875
I18              0.812
I19  0.794
I20  0.499
I23  0.795
I26  0.724
I27  0.711
I28  0.783

                 MR1   MR3   MR2
SS loadings    6.844 2.570 1.497
Proportion Var 0.342 0.129 0.075
Cumulative Var 0.342 0.471 0.546


Standardized loadings (pattern matrix) based upon correlation matrix
      MR1   MR3   MR2   h2   u2 com
I1  -0.05  0.88  0.04 0.74 0.26 1.0
I2   0.02  0.82 -0.07 0.68 0.32 1.0
I3   0.01  0.83 -0.02 0.70 0.30 1.0
I4   0.20  0.58  0.02 0.54 0.46 1.2
I5   0.64  0.12 -0.03 0.52 0.48 1.1
I7   0.68  0.14 -0.08 0.60 0.40 1.1
I8   0.72  0.05  0.04 0.58 0.42 1.0
I10  0.81 -0.06  0.06 0.61 0.39 1.0
I11 -0.03  0.13  0.65 0.46 0.54 1.1
I12  0.16  0.05  0.63 0.48 0.52 1.1
I13  0.16  0.59  0.27 0.64 0.36 1.6
I14  0.61  0.17 -0.09 0.52 0.48 1.2
```

```
I16  0.67  0.02  0.01 0.47 0.53 1.0
I17  0.87 -0.01 -0.06 0.73 0.27 1.0
I18 -0.06 -0.06  0.79 0.60 0.40 1.0
I19  0.80 -0.01  0.07 0.66 0.34 1.0
I20  0.50  0.04  0.06 0.30 0.70 1.0
I23  0.79  0.06 -0.09 0.67 0.33 1.0
I26  0.73 -0.03  0.05 0.51 0.49 1.0
I27  0.72 -0.10  0.08 0.46 0.54 1.1
I28  0.77 -0.06 -0.01 0.54 0.46 1.0

                       MR1  MR3  MR2
SS loadings           7.15 3.23 1.61
Proportion Var        0.34 0.15 0.08
Cumulative Var        0.34 0.49 0.57
Proportion Explained  0.60 0.27 0.13
Cumulative Proportion 0.60 0.87 1.00

 With factor correlations of
     MR1  MR3  MR2
MR1 1.00 0.69 0.18
MR3 0.69 1.00 0.19
MR2 0.18 0.19 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  210  and the objective function was
13.21 with Chi Square of  11861.6
The degrees of freedom for the model are 150  and the objective function was  0.94

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

The harmonic number of observations is  895 with the empirical chi square  302.35
with prob <  2.5e-12
The total number of observations was  907  with Likelihood Chi Square =  842.79
with prob <  6e-97

Tucker Lewis Index of factoring reliability =  0.917
RMSEA index =  0.071  and the 90 % confidence intervals are  0.067 0.076
BIC =  -178.73
Fit based upon off diagonal values = 1
Measures of factor score adequacy
```