

Running head: ASSESSMENT OF EMOTION UNDERSTANDING WITH STEU-B1

Rasch Model analysis of the Situational Test of Emotional Understanding – Brief in a2  
large Portuguese sample3  
4  
5  
6  
7  
8  
9  
10

11Abstract

12  
13  
14The conceptualization of Emotional Intelligence (EI) has a significant impact on the construct’s  
15operationalization and measurement. **Objectives:** The current study analyses the Situational Test  
16of Emotion Understanding – Brief (STEU-B), a test devised to assess the ability to understand  
17emotional experiences. **Method:** This cross-sectional study includes Rasch Model analysis in a  
18sample of 899 participants from the Portuguese general population, between 14 and 72 years old  
19(M = 22.46). **Results:** The measure showed adequate psychometric properties, with good global  
20and item fit, and adequate range of item difficulty. Differential item functioning was explored by  
21gender, level of education and age. Divergent and convergent validity were explored with gender,  
22age, and performance in computerized social and complex cognition tasks. **Conclusions:** The  
23current work suggests the STEU-B is a psychometrically sound test that assesses emotional  
24understanding, and it constitutes an important contribution to the diversification of EI-related  
25measures as well as to research aimed at refining our current understanding of emotional  
26understanding and related abilities.  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47

48*Keywords:* emotional intelligence; emotion understanding; assessment; IRT; Rasch model  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

Emotional intelligence (EI) became an alluring construct to the wider public, as it is believed to encompass personal abilities that can be advantageous in a variety of settings (e.g. intra and interpersonal relationships with family, peers, in schools or workplace environments, etc.). However, most scientific evidence supporting this notion is correlational in nature, consisting of a plethora of studies between EI and several overlapping psychological constructs (e.g. personality, intelligence) and outcomes that, at best, suggest evidence of a small improvement in academic or professional success and quality of life (Côté & Miners, 2006; Gardner & Stough, 2002; Joseph & Newman, 2010; Parker, Summerfeldt, Hogan, & Majeski, 2004; Power & Dalglish, 2007; Schutte, Malouff, Thorsteinsson, Bhullar, & Rooke, 2007; Zeidner, Matthews, & Roberts, 2004). Research also tends to overlook hypotheses related to negative outcomes of high EI skills (Côté, Decelles, McCarthy, Van Kleef, & Hideg, 2011). Furthermore, the concept of EI still raises a considerable debate within the scientific communities over different conceptualization of the construct, along with criticisms and cautionary notes about the pitfalls and overreliance in existing EI assessment abound (Fan, Jackson, Yang, Tang, & Zhang, 2010; Joseph & Newman, 2010; Roberts, Zeidner, & Matthews, 2001b). While some authors regard EI as personality traits or behavioural tendencies, others propose EI is a set of cognitive abilities or competencies (Evans, Hughes, & Steptoe-Warren, 2019; MacCann, Joseph, Newman, & Roberts, 2014; MacCann & Roberts, 2008; Mayer, Roberts, & Barsade, 2008; Mayer & Salovey, 1997; Mayer, Salovey, & Caruso, 2000, 2008). Those theoretical stances have direct implications for measurement of EI and, consequently, on any theoretical formulation of models and empirical evidence involving the construct.

Meta-analytic studies indicate that, despite the long history of theoretical and empirical research in ability EI in the psychological fields, evidence for criterion validity is still very scarce for ability models (Joseph & Newman, 2010). Empirical studies also highlight the scarcity and inconsistency of findings regarding gender and race differences in EI ability, and the need to include community samples and address possible biases in ability testing for different sociodemographic groups (Allen, Weissman, Hellwig, MacCann, & Roberts, 2014; Fan et al., 2010; MacCann & Roberts, 2008; Sharma, 2017). In an examination by Petrides (2011), it was argued that a clearer distinction between the different theoretical proposals is necessary, because many assessment measures present convergence problems and show interrelationships with different constructs (intelligence, effectiveness and declarative knowledge, and other cognitive abilities) (Barchard & Hakstian, 2004; Mayer, Caruso, & Salovey, 1999; Petrides, 2011; Roberts, Zeidner, & Matthews, 2001a). This issue highlighted the need to rethink assessment methods of emotional intelligence components, whether by making them similar to traditional intellectual and cognitive abilities tests, or by creating novel approaches that are more suitable for the measurement of this construct and related personality and cognitive abilities.

We adopt the conceptualization of EI as a set of abilities, similarly to the theoretical model initially proposed by Mayer and colleagues (Mayer & Salovey, 1997). The hierarchical model of Emotional abilities includes skills related to how individuals perceive and express their emotions, how emotions are integrated into other thought processes, the understanding of the complex interplay between emotions and context, and the management or regulation of emotional experiences. While early formulations postulated a 4-branch model, empirical evidence has pointed out to a 3-factor hierarchical structure (perception, understanding and management) as a part of a broader second-stratum ability-EI factor within a general intelligence

structure (Evans et al., 2019; Fan et al., 2010; MacCann et al., 2014). Furthermore, these components of emotional processing skills also fall within the wider domain of social cognition, which is related to, but distinct from, other non-social cognitive abilities (da Motta, Carvalho, Castilho, & Pato, 2019; Frajo-Apor et al., 2017; Lysaker et al., 2014; Phillips, Drevets, Rauch, & Lane, 2003; Schmidt, Mueller, & Roder, 2011; Singer, 2012).

The focus of the current study is the evaluation of emotion understanding, a skill that can contribute to the measurement of the general concept of EI. For years, the most widely used measure to date has been the MSCEIT, a measure that has received considerable criticism regarding its construction, scoring methods yielding inconsistent results, and validity issues (Fan et al., 2010; MacCann & Roberts, 2008; Petrides, 2010; Roberts et al., 2006). In an attempt to increase the diversity of measurement of EI, and to surpass its known shortcomings, MacCann and Roberts (2008) developed the Situational Test of Emotion Understanding (STEU), which presented promising results for the evaluation of the third branch of Emotional Intelligence.

Few psychometrically sound EI measures have been made freely and directly accessible to researchers and clinicians. Although most EI assessment tools are proprietary measures subjected to considerable reproduction and administration fees, the permissions for reproduction and use of the STEU can be easily obtained from the APA PsycTests (see method section). The STEU is available in long (42 items) and brief (19 items) forms (Allen et al., 2014; MacCann & Roberts, 2008). The test items were devised within the Roseman's appraisal theory (Roseman, 2001) and consist of several scenarios that participants can rate, in a multiple-choice format, which emotion is likely to be felt in that given situation. Responses are considered correct or incorrect according to that underlying theoretical model. This method has demonstrated to be a good approach to maximum performance testing of emotional abilities (Austin, 2010; MacCann

& Roberts, 2008) and previous studies suggested emotional understanding ability holds the strongest relationship to cognitive abilities, supporting the idea that emotion understanding involves conceptual knowledge and representations about emotions (Allen et al, 2014; MacCann, 2010; MacCann, et al 2014). While both forms presented good psychometric properties in their initial studies, the STEU presented a slightly lower reliability in its shortest form in comparison to the longer version. This is a common trade-off when retaining the smallest pool of items (less than 50%, in this case) that preserve the largest possible amount of information in a measure (Allen et al., 2014). The use of brief measures, whether in clinical or in the research field, is particularly emphasized due to their cost-effectiveness, especially when working with more vulnerable individuals or avoiding bias due to respondent fatigue or carelessness. In this regard, previous research also points out to the advantages of carrying studies using Item Response Theory (IRT) to explore estimates of ability and differential item functioning according to sociodemographic characteristics, since studies about this measure (and of EI measures in general) in independent samples are still lacking (Allen, Weissman, et al., 2014; Barchard & Hakstian, 2004; Barchard & Russell, 2006; Fan et al., 2010; MacCann & Roberts, 2008; Sharma, 2017). To the best of our knowledge, no psychometric studies were further carried out with the brief version of the STEU. Thus, the goal of the current study is twofold: to translate and adapt the STEU-B to make it available for researchers and practitioners working with Portuguese speaking communities, and to study the psychometric properties of the STEU-B using item response theory and the relationship with more basic EI processes and cognitive abilities. In addition, we expand previous works by studying differential item functioning and suggesting scoring techniques that can more precisely locate an individual's skill level across the Emotional Understanding construct.

## Method

### Participants

A total of 899 participants took part in this study. Three-hundred sixty participants (40%) were males and 539 (60%) were females, with ages between 14 and 72 years old ( $M = 22.46$ ;  $SD = 9.98$ ). Most participants were single ( $N = 748$ ; 83.2%), followed by 125 (13.9%) participants who were married or in a civil union, 21 participants who were divorced (2.3%) and 2 widowed (0.2%). Three participants (0.3%) did not report their civil status. Most participants ( $N = 388$ ; 43.2%) were attending or completed elementary school and 344 (38.3%) were attending or completed mandatory education (high school). The remaining 151 (16.7%) participants had a higher education (college, masters or doctoral degree) or alternative curricula (e.g. professional school), while 16 participants did not report their education level (1.7%).

### Measures

*Situational Test of Emotional Understanding – Brief* (STEU-B) (MacCann & Roberts, 2008, Portuguese version by da Motta, C., Barreto Carvalho, C. & Castilho, P., 2016). The STEU-B is a 19-item version of the STEU (42 items), obtained from IRT analysis performed in the studies by Allen et al. (2014), that provides the maximum amount of information with the least testing time. The items depict different interpersonal scenarios, with different degrees of difficulty, to which participants are asked to choose the more likely emotional response resulting from each situation (e.g. *Clara receives a gift. Clara is most likely to feel?* Response options: a) happy; b) angry; c) frightened; d) bored; e) hungry). Items were scored dichotomously, rated as correct or incorrect according to the key provided by the authors (C. MacCann, personal communication, March 22<sup>nd</sup>, 2016) and available as a supplemental file in the paper by MacCann

& Roberts (2008). The STEU-B is the shortest available ability-based Emotional Intelligence test of emotion understanding, with increased utility for assessment time restrictions and suitable for the general public (Allen et al., 2014).

*Penn Emotion Recognition Task* (da Motta et al., 2019; Gur et al., 2010) is a measure of social cognition that tests ability to decode and identify specific emotions presented in 40 faces (fear, anger, sadness, joy or no emotion), balanced by gender and extreme or mild emotional expressions.

*Short Penn Verbal Reasoning Test* (da Motta et al., 2019; Gur et al., 2010) is a measure of verbal intellectual ability, comprising the 8 best predictors of the full 29-questions version of the test (Gur et al., 2001). It consists of verbal analogies that are answered in a multiple-choice format.

*Short Raven Progressive Matrices* (da Motta et al., 2019; Gur et al., 2010) comprises a computerized version of 9 questions from the 60-question regular Raven test, chosen according to a previous analysis that demonstrated their predictive power of the scores from the 60-questions version (Gur et al., 2001). The task measures non-verbal reasoning for solving each matrix problem.

## Procedures

Prior to data collection, the authors obtained permission to translate, reproduce and use the Situational Test of Emotion Understanding – Brief (see Measures section) from the authors and APA<sup>1</sup>. We followed several guidelines for adequate cross-cultural adaptation of psychological assessment measures (Beaton, Bombardier, Guillemin, & Ferraz, 2000; Sousa & Rojjanasrirat, 2011). The only change in the content was made in some of the first names of the

---

<sup>1</sup> The STEU-B and STEM-B are APA-copyrighted material. Permissions to reproduce or distribute the STEU-B and STEM-B translation require the written permission from the APA ([permissions@apa.org](mailto:permissions@apa.org)).

persons described in STEU scenarios, which were changed to names or spellings more familiar to the Portuguese native-speakers (e.g. Susan was replaced by Suzana). The first translation of the contents to Portuguese was carried out by a bilingual psychologist (first author). The back translation was performed by a U.S resident bilingual teacher that was blind to the study goals. No significant deviations from the original version of the test were found between the original English version and back-translated version, and the final Portuguese translation was revised by a senior psychologist. Changes made at this stage referred to grammar/punctuation, or deciding the best term whenever an English word could be translated in more than one way in Portuguese. Prior to its administration to a wider sample, the Portuguese STEU-B was administered to 5 adult participants from the general population. In the pilot administration, participants were asked to carefully read all the scenarios and response options before responding, and to mark any items in which they felt something was wrong, difficult to read or had words they did not understand. After the test was completed, the researcher asked each participant about any issues found in the STEU-B and participants did not report difficulties regarding items' clarity or comprehension.

A convenience sample of 899 participants was administered a research protocol including the STEU-B as a part of a wider research project. Participants were adolescents and adults from the Portuguese mainland and Azores islands communities. All participants were informed about the research goals, warranting the anonymity and the voluntary character of participation, and only participants who provided their written informed consent were administered the measures included in the research protocol. Adult participants were recruited through non-probabilistic sampling method (e.g., the authors' contact networks, advertisements in local institutions, newspapers and other media). Underage participants were contacted in local schools and a signed informed consent was obtained from their legal representatives prior to participation.



### Statistical Analyses

The statistical procedures were computed using WINSTEPS Rasch Analysis (version 3.93, SWREG Inc., 2017), MPLUS (Muthén & Muthén, 2010) and IBM SPSS Statistics (version 23 for Microsoft Windows, IBM Inc. Armonk, NY) software.

The large and diverse sample size in the current study warrants the robustness of statistical analyses using the Item Response and Classical Test theories. The STEU-B was analysed through Rasch Model analysis (RM; Rasch, 1961), as responses differed from item to item in terms of content, and so that items could be evaluated individually and independently. The one-dimensional structure of the STEU-B was firstly assessed with a Confirmatory Factor Analysis (CFA) and assessed through Principal Component Analysis of Residuals (PCAR), or contrasts of standardized Residuals variance (Aryadoust & Raquel, 2019). *Eigenvalues* of unexplained variance obtained on the second contrast should not exceed 2 (Linacre, 2011; Raîche, 2005), otherwise it indicates multidimensionality problems. The local independence, another assumption of RM, was assured by calculating residual correlations between items and Yen's Q3 was used to identify possible dependent pairs of items. Although no critical values are defined for Yen's Q3, the largest observed correlation of residuals between pairs of items should not be over the absolute value of .20 (Christensen, Makransky, & Horton, 2017; Linacre, 2011).

Parameters of persons and items in RM are transformed on a unit called "measure" (represented by a  $\theta$ , theta) that is distributed along a continuum. The units of measure of  $\theta$  are called logits (log-odds units), a scale with theoretical ranges being  $\pm$  infinite, but typical amplitude ranging between  $\pm 5$  (Prieto & Velasco, 2006), and 0 being the average difficulty point set for the measure.

The Infit Mean Square (Infit MNSQ) and Outfit Mean Square (Outfit MNSQ) was calculated to evaluate the fit of the data to the model. The Infit is a standardized information (weighted mean square) that provides information regarding items and possible structural problems (Baker, 2001; Prieto & Velasco, 2006). Infit statistics values between 0.5 and 1.5 are productive for measurement, while values greater than 2.0 can degrade the measurement; values between 1.5 and 2.0 are regarded as unproductive for measurement construction; and values smaller than 0.5 are less productive for measurement, but not degrading (Linacre, 2011). The Outfit is an unweighted mean square and is sensitive to outliers (Linacre, 2011). Both indices are produced for items and persons.

To provide initial evidence of the validity of the STEU-B, we assessed possible item bias through Differential Item Functioning (DIF) analyses. We inspected for possible changes in item's difficulty parameters across gender, educational level and age while keeping ability constant. It is expected that the items do not present differences in difficulty (e.g. favours a group over another) between male and female participants, participant's educational level or age group. Whenever significant DIF is found with Rasch-Welch and Mantel-Haenszel tests, DIF that yield a logit difference below .43 are considered negligible, values above .44 are considered slight to moderate, and large difference in item functioning exists across groups when above .64 logits (Zwick, Thayer, & Lewis, 1999).

Lastly, we complemented these analyses with Classical Test Theory practices by calculating convergent validity and exploring gender differences in STEU-B scores, and the relationship between STEU-B scores with age and performance in computerized neurocognitive tests. The scores from the computerized tasks were obtained from a sample of 170 adult participants and calculated from the total number of correct responses in each task (accuracy

scores). We reported Pearson product-moment correlations with bootstrap resampling method set to 2000 samples and 95% bias-corrected Confidence Intervals.

## Results

### Dimensionality

Results from the unidimensional CFA model yielded adequate fit to the data:  $\chi^2_{(152)} = 328.975, p < .001$ ; CFI = .92; TLI = .91; RMSEA = .036  $P[\text{rmsea} < .05] = 1$ . On an item-level analysis, item loadings ranged between -.16 and .75, the items 3, 4, 5, 16 and 18 presented individual loadings below .25. A closer inspection of the items presenting lower loadings revealed those items presented the highest percentage of incorrect response (over 55% receiving 0 points in each item). The value obtained through PCAR analysis was 1.43, which also provided support for a unidimensional model. Because Rasch model analyses provide information on item difficulty and no multidimensionality problems were identified in this phase, items with lower loadings were kept in the model for further analyses.

Low residual correlations were observed between items, with values ranging from  $r = -.18$  to  $r = -.13$ . Yen's Q3 tests yielded weak positive correlations for all items ( $r = -.19$  to  $r = -.13$ ), indicating local independence for all pairs of items.

### Global fit statistics

The model tested has involved all participants and items from the STEU-B. Persons' and items' global fit measures are presented in Table 1. Most items presented adequate adjustments, with average Infit and Outfit values of .98 ( $SD = 0.09$ ) and 1.22 ( $SD = .66$ ), respectively. The maximum value of the Outfit was of 3.61, which suggests the presence of outliers or, at least, one

item with poor adjustment. The amplitude of the measure for the items ranged from -1.58 to 2.41 logits and the measure's standard error was low (between .07 and .11,  $M = 0.08$ ;  $SD = 0.01$ ). Person fit showed appropriate to *Infit* ( $M = .98$ ,  $SD = 0.25$ ) and *Outfit* ( $M = 1.19$ ,  $SD = 1.11$ ) average values. The maximum values of outfit also inform about the existence of participants with results that did not fit the model. The inspection of extreme infit and outfit values revealed that about 32 participants presented response patterns diagnosed matching "careless" or "lucky guessing" (Linacre & Wright, 1994). All participants that presented this pattern were adolescents (participants under 18 years old) and were excluded from analysis. Because these response patterns might be a possible cause of measure distortions observed in item's fit statistics, the model was analysed excluding the 32 participants identified as presenting problematic response patterns.

(insert table 1)

Global fit statistics presented in table 2 showed that the current set of items has the necessary conditions to produce an assessment tool with adequate metric properties. Item's fit statistic now falls within adequate values of infit and outfit, without any indicators of measure degradation within the model (*Infit* and *Outfit* < 2, table 3). Reliability and separation values for person was .60 and 1.23 (lower bound), indicating a lower ability to discriminate participants in the sample according to their level of performance, while the item reliability and separation were .99 and 14.07 (lower bound), suggesting a good item difficulty hierarchy. The Cronbach alpha (KR-20) for the scale was .63.

(insert table 2)

### Items-person map

The distribution of persons and items across the measure ( $\theta$ ) is represented in figure 1. Most items align vertically across the logits scale, with only 3 pairs of items appearing parallel. In other words, the items in parallel positions represent a similar degree of ability (emotional understanding), and these items could be excluded to make the measure shorter. However, no item was excluded from the analysis because those items did not present any fit problems and it is preferable to keep the STEU-B identical across all versions. The items also referred to distinct interpersonal scenarios that may be relevant to the theoretical construct at hand and screening for careless/random response patterns (e.g. a participant with a certain degree of emotional understanding ability has the same probability of choosing the correct response irrespective of the contexts in which a situation takes place).

(insert figure 1)

The average  $\theta$  for persons was .06 logits, as the  $\theta$  average for the items is zero by convention. Most participants fall within a range of 4 logits (between -2 and 2), with items 3 and 4 presenting less frequent rate of correct responses. This suggests items can evaluate higher capacity of emotion understanding than what was observed in most participants. The items measuring the highest degrees of emotional understanding are, thus, items 3 and 4 (placed near point 3 of the logits scale), and the items measuring the lesser degree are items 6 and 13 (placed near point -2 of the logits scale). Finally, each item's statistics of STEU-B present adequate values of Infit and Outfit values, low standard errors and adequate point-bisserial correlations to the total measure, as presented in table 3.

(insert table 3)

### Differential Item Functioning

To test for differences in item functioning, we carried out Differential Item Functioning (DIF) analysis of items across groups according to gender (males vs. females), age (participants with more vs. less than 18 years old) and years of education (mandatory education or below vs. higher education). As shown in figure 2 (A), no statistically significant differences in item difficulty was found in pairwise comparisons in males and female participants, except for item 17, which reached the significant threshold and presented a .42 logits favouring males, a DIF value within a negligible threshold according to Zwick, Thayer & Lewis (1999). Regarding educational level, 5 items presented statistically significant differences in measures in pairwise comparisons of participants with mandatory education or less, or higher education. As shown in fig. 2 (B), 3 items favoured those with higher education levels, while 2 favoured the group with less years of education. The magnitude of this difference was considered moderate to large, as the 5 items DIF contrasts had values above .64 logits (or below -.64 logits, when items favoured participants with less years of education). Finally, when comparing groups regarding age (below or above 18 years old), nearly half the items presented large differences in item's difficulty measures, as presented in fig. 2 (C) and table 4. A more thorough analysis of the plot and the table show that items favouring youths were located among the easiest and medium difficulty items on the DIF measure plot, and the more difficult items tend to favour the older participants. Differential Group Functioning analysis was non-significant, indicating that observed differences in item difficulty between age groups did not impact the total score of the STEU-B measure.

### **Convergent and discriminant validity**

To assess convergent and divergent validity, we explored the relationships between test scores and age and sex differences. A positive, but weak statistically significant correlation, was found with age, indicating older participants achieve higher scores in the STEU-B ( $r = .222, p < .001, 95\% \text{ CI } [.166, .277]$ ). Females showed a statistically higher average score ( $M = 50.69; SD = 9.81$ ) in comparison to male participants ( $M = 48.46; SD = 10.47; t_{(897)} = -3.241, p = .001$ ), but the gender effect was small (Cohen's  $d = .22$ ).

Additionally, a sample of 170 participants responded to the STEU-B and the measures of Emotion Recognition, Verbal Reasoning and Raven's Matrices computerized tasks that recorded accuracy (number of correct responses), as a collateral source of information. Correlation analysis showed a significant association with the accuracy in measures of complex cognition assessed by Verbal Reasoning task ( $r = .370, p = .000, 95\% \text{ CI } [.218, .498]$ ), with the Emotion Recognition task ( $r = .259, p = .001, 95\% \text{ CI } [.049, .438]$ ) and Raven's Progressive Matrices test ( $r = .338, p = .000, 95\% \text{ CI } [.186, .480]$ ). Results indicate the convergence of scores involving more sophisticated cognitive processes required to appraise an emotion-eliciting scenario, and a weaker correlation with more elementary emotional processes, as the ability to recognize emotional states from facial stimuli, and non-verbal abstract reasoning.

### **Scoring**

To more accurately take into account the differences in item difficulty and avoid several psychometric issues that arise from summing raw correct responses, test administrators are encouraged to convert raw test scores into logits or to a scaled linear score. Table 5 presents a score conversion to both logits and a linear scale (ranging from 0-100, in order to resemble

percentages), and indicates the respective percentiles to more precisely locate the respondents actual measure and level of performance (Boone, 2016). For example, a respondent that scores 12 correct responses should be attributed a score of .64 logits or 56 points, and his/her performance is slightly above the 70<sup>th</sup> percentile. In the current sample, 138 participants (15.9%) have scored 12 points and almost 80% of participants have 12 or less correct responses.

(insert table 5)

### Discussion

Emotion understanding is a skill of relative complexity, which is why it is considered a component that occupies a higher branch in hierarchical Emotional Intelligence (EI) ability models. It encompasses the interplay between cognitive and non-cognitive abilities to allow the understanding of the complex interplay between emotions and context. Notwithstanding the extensive theoretical debate and empirical research on models of EI ability and the importance of its outcomes to several fields of research and practice (e.g. occupational, educational, clinical, etc.), the issues related with assessment and measurement validity of EI abilities are still unsettled, and new measures and assessment paradigms have arisen in the past decade to address the existing methodological gaps (MacCann, Pearce, & Roberts, 2011; MacCann & Roberts, 2008; Roberts, Schulze, & MacCann, 2008).

The current work sought to provide data on the psychometric properties of the brief version of the Situational Test of Emotion Understanding, a test devised to assess the ability to understand contextual cues that can elicit different emotional responses. We used Rasch Model (Item Response Theory), a suitable approach to ability measurement in dichotomous scales (e.g. 0 = incorrect/ 1 = correct), as it allows to create a measurement unit that considers the item's difficulty within the same construct continuum and provides a qualitative assessment of a scale



relevant for construct validity. Rasch model is a useful approach to psychometrics because it conforms to measurement theory rather than focusing on a measure's performance in a particular sample. In the field of psychology, it offers additional advantages in terms of generalizability of measurement models to distinct samples and populations, and to the identification of extreme or unexpected individual responses and item particularities that are unavailable to Classical Theory testing (Fox & Jones, 1998).

While studies on the reliability and (uni)dimensionality of the scale provide statistical validity, differential item functioning was used to control possible item bias toward specific groups, such as gender, educational level and age. The absence of bias in item difficulty between males and females provides evidence for the application of the STEU-B to the wider public. After ascertaining the invariance of the model, females have scored significantly higher than males, similar to previous studies using EI ability measures (e.g. Brackett & Mayer, 2003). This is possibly due to females being more stimulated to develop certain social skills based on cultural and/or social pressures, although differences in emotional processing skills tend to fade in some contexts (see Brody & Hall, 2010). However, no sufficient evidence has been gathered regarding gender differences in emotion understanding ability, specifically. The differences in item functioning were more prominent regarding age, and findings demonstrated that some items could be somewhat biased for older participants, but others to younger participants. Nevertheless, DIF cancellation was observed in the DIF analysis regarding possible bias according to education or age - items favouring one group or the other were balanced and did not compromise the overall test score (Wyse, 2013). Irrespective of academic level, older participants can be ahead in terms of developmental processes relevant to social skills (e.g. complete brain maturation and development of important cognitive functions underlying emotion understanding).

Furthermore, due to the relationship between Emotion Understanding and crystallized intelligence, which tends to increase over the lifespan, it is possible that this ability may also increase with age (Phillips, MacLean, & Allen, 2002). Older participants are expected to have accumulated experiences and knowledge over their lifespan that may be fundamental to the development of the capacity to more precisely estimate the emotions that can arise from a wide range of interpersonal contexts (Kafetsios, 2004, Sharma, 2017). However, when holding person's abilities constant, it is important to ensure the measure is invariant across participants' characteristics. This aspect can be particularly relevant when devising new items or instruments and when assessing emotion understanding ability in groups with a wide age range.

The reliability (KR-20) obtained in the current study was only adequate, but higher than those of the studies of the 42-item version (MacCann & Roberts, 2008) and equal to those of the brief version by Allen et al. (2014). Lower reliability is expected in RM analyses because of local independence of items, in other words, the responses are not similar, nor depend or correlate with each other. The more typical reliability indices (e.g. Cronbach's alpha, KR-20) are prone to be artificially inflated by local dependence and these same indices can be also affected by the number of alternatives in multiple-response forms (Jensen, 1980). The possibility of studying the item's properties and the ability range of respondents is an advantage of IRT over Classical Test Theories. According to Linacre (2011), a Rasch reliability above .5 can suffice to distinguish at least 2 levels of performance (e.g. high/low performers on a given measure), which is satisfactory for a brief evaluation of emotion understanding ability. The item map showed that the items could be precisely located across the latent variable in terms of difficulty hierarchy, but some gaps were wider in the top half (more difficult items). It is possible that the lower person reliability and separation values found in the current study be due to the reduced number of items

or the sample presenting a narrower ability range, as it included only participants from the general population and no other extreme groups (e.g. clinical samples, participants with deficits in cortico-limbic functions). The ceiling effects found in the original studies of the STEU-B (Allen et al., 2014) and a similar limitation found in the study of the longer version of the STEU, which included a sample with a higher average education and age (MacCann & Roberts, 2008), were not found in the current study. For this reason, the former explanation - the lower person reliability and lower separation values could be simply due to the small pool of items included in the brief version - should be more likely. Hypothetically, the longer version (more items) could be preferable for studies in which there is the need to further discriminate individuals by more nuanced levels of performance, as the increased number of items may increase the sensitivity of the STEU for more intermediate levels of ability. Conversely, the shorter version would be more advisable to avoid participant fatigue, when working with more vulnerable individuals or in studies in which more rough distinction between low and high ability testing would suffice (Allen et al., 2014; Boone, 2016). Future research regarding this measure should explore how the longer and shorter forms are able to discriminate different levels of performance in order to clarify this hypothesis.

Findings regarding convergent and divergent validity showed that emotion understanding correlated positively with verbal reasoning, emotion recognition and general intelligence tests. The moderate and positive association with verbal ability is comprehensible to the extent that the ability to comprehend (e.g. read and/or interpret the descriptions of a scenario) and correctly provide a verbal label to an emotion can be an ability associated with emotion understanding, especially when tests are presented in a written format (MacCann et al., 2014). While it would be expected that EI abilities converge strongly to each other and less with cognitive abilities,

findings are congruous with studies demonstrating the associations between emotion understanding and emotion perception can be significantly lower when tests consist of visual stimuli in comparison to other types of stimuli (Elfenbein & MacCann, 2017). Previous findings also showed emotional understanding had stronger relationships with cognitive abilities in comparison to other EI branches, and these relationships were stronger for crystallized intelligence in comparison with fluid intelligence, providing support to the idea that emotion understanding involves conceptual knowledge and representations about emotions (Allen et al, 2014; MacCann, 2010; MacCann, et al 2014). Nevertheless, studies need to further clarify how emotion understanding relates to other cognitive abilities within existing intelligence frameworks. The sample used in these analyses was small and bias-corrected confidence intervals had a relatively wide range, and future studies should aim to study further the STEU-B convergence, preferably in larger samples for more consistent estimates, and explore the measure's temporal stability.

Overall, the 19-items provide a robust assessment of the emotional understanding construct, and findings are indicative of a substantial equivalence to the original version of the STEU-B (Allen et al., 2014; MacCann & Roberts, 2008). The availability of a shorter form of the measure with a simple scoring procedure (correct/incorrect) is a particularly useful measure when working with specific populations (e.g. clinical samples) or when researchers face assessment time constraints. Test scores diverged from scores in other relevant cognitive processing abilities. Thus, it provides an important contribution to the assessment of EI abilities, as the field can benefit from the diversification of assessment measures and methodologies that can improve the operationalization of the EI-related constructs. The use of novel tests with demonstrated construct validity can fulfil the existing methodological gaps left by the overuse of

the MSCEIT and overcome some difficulties related with the access to commercial measures for EI research. Furthermore, because the STEU-B focuses specifically on the Emotion Understanding construct, it can provide important insights about this EI component and research aimed at refining existing models of EI.

Increased performance in emotional understanding (or the ability to adequately elicit/predict affect and emotional states) is a higher-order process that requires both social and cognitive processing abilities to collect information and form global appraisals from social, contextual and/or motivational cues. Ability-based tools can help further clarify the relationship between specific higher-order EI skills and more complex cognitive abilities, as well as be used in efficacy studies of programs or interventions that promote different emotional skills and more adjusted behaviour (e.g. Cognitive Behavioural Therapy, Mindfulness-based cognitive therapies, Social-Emotional Learning programs, interventions for patient's deficits in empathy, etc.; Dattilio, Kazantzis, Reinecke, & Freeman, 2009; Liotti & Gilbert, 2011; Wölwer et al., 2005). Furthermore, the use of more robust assessment tools of EI ability may be beneficial to managers, educators and practitioners acting in several settings (e.g. schools, workplace, healthcare), as well as researchers carrying out empirical studies with cross-sectional and longitudinal designs in a large variety of areas, that can tackle the potentialities and controversies around Emotional Intelligence.

### Conclusion

The Portuguese version of the STEU-B is a novel and cost-effective measure available for professionals working with Portuguese-speaking communities. The dissemination of this tool is of added-value, taking into consideration the high number of Portuguese native-speakers worldwide (Lewis, Simons, & Fennig, 2016), and the possibility of increasing the diversity of

assessment measures to be used in cross-cultural studies or empirical research on EI and social cognition domain. On the other hand, this aids the refinement of existing models of EI and may be a useful tool to clarify the interdependence between social cognition, cognitive ability and related constructs in future studies. Finally, it should be highlighted that the use of specific tests of emotional skills, including emotion understanding, can help devise more targeted evaluation and intervention strategies in clinical, occupational or educational settings, and identify the mechanisms related to adjusted interpersonal functioning and psychological well-being.

### **Compliance with Ethical Standards**

**Conflict of Interest:** The authors declare that they have no conflict of interest.

**Ethical approval:** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed Consent:** Informed consent was obtained from all individual participants, or their legal representatives, included in the study.

**Data Availability Statement:** The materials and data that support the findings of this study are available from the corresponding author upon request.

## References

- Allen, V. D., Weissman, A., Hellwig, S., MacCann, C., & Roberts, R. D. (2014). Development of the situational test of emotional understanding – brief (STEU-B) using item response theory. *Personality and Individual Differences*, 65, 3–7. <https://doi.org/10.1016/j.paid.2014.01.051>
- Aryadoust, V., & Raquel, M. (2019). Quantitative Data Analysis for Language Assessment Volume I. In V. Aryadoust & M. Raquel (Eds.), Quantitative Data Analysis for Language Assessment Volume I (1st ed.). <https://doi.org/10.4324/9781315187815>
- Baker, F. B. (2001). *The Basics of Item Response Theory*. Washington: ERIC Clearinghouse on Assessment and Evaluation. <https://doi.org/10.1111/j.1365-2362.2010.02362.x>
- Barchard, K. A., & Hakstian, A. R. (2004). The nature and measurement of emotional intelligence abilities: Basic dimensions and their relationships with other cognitive ability and personality variables. *Educational and Psychological Measurement*, 64(3), 437–462. <https://doi.org/10.1177/0013164403261762>
- Barchard, K. A., & Russell, J. A. (2006). Bias in consensus scoring, with examples from ability emotional intelligence tests. *Psicothema*, 18(SUPPL.1), 49–54.
- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine*, 25(24), 3186–3191. <https://doi.org/10.1097/00007632-200012150-00014>
- Boone, W. J. (2016). Rasch analysis for instrument development: Why,when,and how? *CBE Life Sciences Education*, 15(4). <https://doi.org/10.1187/cbe.16-04-0148>
- Brackett, M. A., & Mayer, J. D. (2003). Convergent, discriminant, and incremental validity of competing measures of emotional intelligence. *Personality and Social Psychology Bulletin*,

29(9), 1147–1158. <https://doi.org/10.1177/0146167203254596>

Brody, L. R., & Hall, J. A. (2010). Gender, Emotion, and Socialization. In J. Chrisler & D. McCreary (Eds.), *Handbook of Gender Research in Psychology* (pp. 429–454).  
[https://doi.org/10.1007/978-1-4419-1465-1\\_21](https://doi.org/10.1007/978-1-4419-1465-1_21)

Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Applied Psychological Measurement*, 41(3), 178–194.  
<https://doi.org/10.1177/0146621616677520>

Côté, S., Decelles, K. A., McCarthy, J. M., Van Kleef, G. A., & Hideg, I. (2011). The Jekyll and Hyde of emotional intelligence: emotion-regulation knowledge facilitates both prosocial and interpersonally deviant behavior. *Psychological Science*, 22(8), 1073–1080.  
<https://doi.org/10.1177/0956797611416251>

Côté, S., & Miners, C. T. H. (2006). Emotional Intelligence, Cognitive Intelligence, and Job Performance. *Administrative Science Quarterly*, 51(1), 1–28.  
<https://doi.org/10.2189/asqu.51.1.1>

da Motta, C., Barreto Carvalho, C., & Castilho, P. (2016). *Situational Test for Emotional Understanding - Brief [Portuguese Version, Unpublished instrument]*.

da Motta, C., Carvalho, C. B., Castilho, P., & Pato, M. T. (2019). Assessment of neurocognitive function and social cognition with computerized batteries: Psychometric properties of the Portuguese PennCNB in healthy controls. *Current Psychology*.  
<https://doi.org/10.1007/s12144-019-00419-2>

Dattilio, F., Kazantzis, N., Reinecke, M., & Freeman, A. (2009). *Cognitive and behavioral theories in clinical practice*. New York: The Guilford Press.

Elfenbein, H. A., & MacCann, C. (2017). A closer look at ability emotional intelligence (EI):



- What are its component parts, and how do they relate to each other? *Social and Personality Psychology Compass*, 11(7), e12324. <https://doi.org/10.1111/spc3.12324>
- Evans, T. R., Hughes, D. J., & Steptoe-Warren, G. (2019). A Conceptual Replication of Emotional Intelligence as a Second-Stratum Factor of Intelligence. *Emotion*, (April). <https://doi.org/10.1037/emo0000569>
- Fan, H., Jackson, T., Yang, X., Tang, W., & Zhang, J. (2010). The factor structure of the Mayer-Salovey-Caruso Emotional Intelligence Test V 2.0 (MSCEIT): A meta-analytic structural equation modeling approach. *Personality and Individual Differences*, 48(7), 781–785. <https://doi.org/10.1016/j.paid.2010.02.004>
- Fox, C. M., & Jones, J. A. (1998). Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology*, 45(1), 30–45. <https://doi.org/10.1037/0022-0167.45.1.30>
- Frajo-Apor, B., Kemmler, G., Pardeller, S., Plass, T., Mühlbacher, M., Welte, A.-S., ... Hofer, A. (2017). Emotional intelligence and non-social cognition in schizophrenia and bipolar I disorder. *Psychological Medicine*, 47(01), 35–42. <https://doi.org/10.1017/S0033291716002324>
- Gardner, L., & Stough, C. (2002). Examining the relationship between leadership and emotional intelligence in senior level managers. *Leadership & Organization Development Journal*, 23(2), 68–78. <https://doi.org/10.1108/01437730210419198>
- Gur, R. C., Gur, R. E., Ragland, J. D., Moberg, P. J., Turner, T. H., Bilker, W. B., ... Siegel, S. J. (2001). Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology*, 25(5), 766–776. [https://doi.org/10.1016/S0893-133X\(01\)00278-0](https://doi.org/10.1016/S0893-133X(01)00278-0)
- Gur, R. C., Richard, J., Hughett, P., Calkins, M. E., Macy, L., Bilker, W. B., ... Gur, R. E. (2010).

- A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: Standardization and initial construct validation. *Journal of Neuroscience Methods*, 187(2), 254–262. <https://doi.org/10.1016/j.jneumeth.2009.11.017>
- Jensen, A. R. (1980). *Bias in Mental Testing*. New York: Free Press.
- Joseph, D. L., & Newman, D. A. (2010). Emotional Intelligence: An Integrative Meta-Analysis and Cascading Model. *Journal of Applied Psychology*, 95(1), 54–78. <https://doi.org/10.1037/a0017286>
- Kafetsios, K. (2004). Attachment and emotional intelligence abilities across the life course. *Personality and Individual Differences*, 37(1), 129–145. <https://doi.org/10.1016/j.paid.2003.08.006>
- Lewis, P., Simons, G., & Fennig, C. (2016). Summary by language size. In P. Lewis, G. Simons, & C. Fennig (Eds.), *Ethnologue: Languages of the World* (19th ed.). Retrieved from <http://www.ethnologue.com/statistics/size>
- Linacre, J. M. (2011). *A User's Guide to Winsteps/ Ministeps: Rasch-Model Computer Programs*. <https://doi.org/ISBN 0-941938-03-4>
- Linacre, J. M., & Wright, B. D. (1994). (Dichotomous Mean-square) Chi-square fit statistics. *Rasch Measurement Transactions*, 8(2), 360. Retrieved from <https://www.rasch.org/rmt/rmt82a.htm>
- Liotti, G., & Gilbert, P. (2011). Mentalizing, motivation, and social mentalities: theoretical considerations and implications for psychotherapy. *Psychology and Psychotherapy*, 84(1), 9–25; <https://doi.org/10.1348/147608310X520094>
- Lysaker, P. H., Leonhardt, B. L., Brüne, M., Buck, K. D., James, A., Vohs, J., ... Dimaggio, G. (2014). Capacities for theory of mind, metacognition, and neurocognitive function are

independently related to emotional recognition in schizophrenia. *Psychiatry Research*, 219(1), 79–85. <https://doi.org/10.1016/j.psychres.2014.05.004>

MacCann, C. (2010). Further examination of emotional intelligence as a standard intelligence: A latent variable analysis of fluid intelligence, crystallized intelligence, and emotional intelligence. *Personality and Individual Differences*, 49(5), 490–496. <https://doi.org/10.1016/j.paid.2010.05.010>

MacCann, C., Joseph, D. L., Newman, D. A., & Roberts, R. D. (2014). Emotional intelligence is a second-stratum factor of intelligence: Evidence from hierarchical and bifactor models. *Emotion*, 14(2), 358–374. <https://doi.org/10.1037/a0034755>

MacCann, C., Pearce, N., & Roberts, R. D. (2011). Emotional intelligence as assessed by situational judgment and emotion recognition tests: Building the Nomological net. *Psihologijske Teme*, 20(3), 393–412.

MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: theory and data. *Emotion*, 8(4), 540–551. <https://doi.org/10.1037/a0012746>

Mayer, J. D., Caruso, D. R., & Salovey, P. (1999). Emotional intelligence meets traditional standards for an intelligence. *Intelligence*, 27(4), 267–298. [https://doi.org/10.1016/S0160-2896\(99\)00016-1](https://doi.org/10.1016/S0160-2896(99)00016-1)

Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: emotional intelligence. *Annual Review of Psychology*, 59, 507–536. <https://doi.org/10.1146/annurev.psych.59.103006.093646>

Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & J. Sluyter (Eds.), *Emotional development and emotional intelligence: Implications for Educators* (pp.

- 3–34). New York: Harper Collins. Mayer, J. D., Salovey, P., & Caruso, D. (2000). Models of emotional intelligence. *Handbook of Intelligence*, pp. 396–420.  
<https://doi.org/10.1080/14766080903069364>
- Mayer, J. D., Salovey, P., & Caruso, D. (2000). Models of emotional intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 396–420).  
<https://doi.org/10.1080/14766080903069364>
- Mayer, J. D., Salovey, P., & Caruso, D. R. (2008). Emotional intelligence: new ability or eclectic traits? *The American Psychologist*, 63(6), 503–517. <https://doi.org/10.1037/0003-066X.63.6.503>
- Muthén, L. K., & Muthén, B. O. (2010). Mplus User's Guide (6th Edition; L. K. Muthén & B. O. Muthén, Eds.). Retrieved from <https://www.statmodel.com/download/usersguide/MplusUsersGuidev6.pdf>
- Parker, J. D. A., Summerfeldt, L. J., Hogan, M. J., & Majeski, S. A. (2004). Emotional intelligence and academic success: examining the transition from high school to university. *Personality and Individual Differences*, 36(1), 163–172. [https://doi.org/10.1016/S0191-8869\(03\)00076-X](https://doi.org/10.1016/S0191-8869(03)00076-X)
- Petrides, K. V. (2010). Trait emotional intelligence theory. *Industrial and Organizational Psychology*, 3, 136–139. <https://doi.org/10.1111/j.1754-9434.2010.01213.x>
- Petrides, K. V. (2013). Ability and Trait Emotional Intelligence. In T. Chamorro - Premuzic, S. von Stumm, & A. Furnham (Eds.), *The Wiley-Blackwell Handbook of Individual Differences* (pp. 656 – 678). <https://doi.org/10.1002/9781444343120.ch25>
- Phillips, L. H., MacLean, R. D. J., & Allen, R. (2002). Age and the understanding of emotions: Neuropsychological and sociocognitive perspectives. *Journals of Gerontology - Series B*

- Psychological Sciences and Social Sciences*, 57(6), 526–530.  
<https://doi.org/10.1093/geronb/57.6.P526>
- Phillips, M. L., Drevets, W. C., Rauch, S. L., & Lane, R. (2003). Neurobiology of emotion perception II: Implications for major psychiatric disorders. *Biological Psychiatry*, 54(5), 515–528. [https://doi.org/10.1016/S0006-3223\(03\)00171-9](https://doi.org/10.1016/S0006-3223(03)00171-9)
- Power, M., & Dalgleish, T. (2007). Cognition and Emotion (Second Edition; M. Power & T. Dalgleish, Eds.). London: Psychology Press. <https://doi.org/10.4324/9780203934487>
- Prieto, G., & Velasco, A. D. (2006). Visualização espacial, raciocínio indutivo e rendimento acadêmico em desenho técnico. *Psicologia Escolar e Educacional*, 10(1), 11–20.  
<https://doi.org/10.1590/S1413-85572006000100002>
- Raîche, G. (2005). Critical Eigenvalue Sizes (Variances) in Standardized Residual Principal Components Analysis (PCA). Retrieved May 8, 2017, from Rasch Measurement Transactions website: <http://www.rasch.org/rmt/rmt191h.htm>
- Rasch, G. (1961). Probabilistic Models for Some Intelligence and Attainment Tests. *Information and Control*, 4(4), 382. [https://doi.org/10.1016/S0019-9958\(61\)80061-2](https://doi.org/10.1016/S0019-9958(61)80061-2)
- Roberts, R. D., Schulze, R., & MacCann, C. (2008). The measurement of emotional intelligence: A decade of progress? In *The SAGE Handbook of Personality Theory and Assessment: Volume 2 - Personality Measurement and Testing* (pp. 461–482).  
<https://doi.org/10.4135/9781849200479.n22>
- Roberts, R. D., Schulze, R., O'Brien, K., MacCann, C., Reid, J., & Maul, A. (2006). Exploring the validity of the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) with established emotions measures. *Emotion*, 6(4), 663–669. <https://doi.org/10.1037/1528-3542.6.4.663>

- Roberts, R. D., Zeidner, M., & Matthews, G. (2001a). Does Emotional Intelligence Meet Traditional Standards for an Intelligence? Some New Data and Conclusions. *Emotion, 1*(3), 196–231. <https://doi.org/10.1037/1528-3542.1.3.196>
- Roberts, R. D., Zeidner, M., & Matthews, G. (2001b). Slow down, you move too fast: Emotional intelligence remains an “elusive” intelligence. *Emotion, 1*(3), 265–275. <https://doi.org/10.1037/1528-3542.1.3.265>
- Roseman, I. (2001). A model of appraisal in the emotion system: Integrating theory, research, and applications. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 68–91). Oxford, England: Oxford University Press.
- Schmidt, S. J., Mueller, D. R., & Roder, V. (2011). Social cognition as a mediator variable between neurocognition and functional outcome in schizophrenia: Empirical review and new results by structural equation modeling. *Schizophrenia Bulletin, 37*(SUPPL. 2), S41–54. <https://doi.org/10.1093/schbul/sbr079>
- Schutte, N. S., Malouff, J. M., Thorsteinsson, E. B., Bhullar, N., & Rooke, S. E. (2007). A meta-analytic investigation of the relationship between emotional intelligence and health. *Personality and Individual Differences, 42*(6), 921–933. <https://doi.org/10.1016/j.paid.2006.09.003>
- Sharma, D. (2017). Impact of Age on Emotional Intelligence and Its Components. *International Journal of Research and Innovation in Social Science, 1*(1), 13–20.
- Singer, T. (2012). The past, present and future of social neuroscience: a European perspective. *NeuroImage, 61*(2), 437–449. <https://doi.org/10.1016/j.neuroimage.2012.01.109>
- Sousa, V. D., & Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline.

- Journal of Evaluation in Clinical Practice*, 17(2), 268–274. <https://doi.org/10.1111/j.1365-2753.2010.01434.x>
- Wölwer, W., Frommann, N., Halfmann, S., Piaszek, A., Streit, M., & Gaebel, W. (2005). Remediation of impairments in facial affect recognition in schizophrenia: Efficacy and specificity of a new training program. *Schizophrenia Research*, 80(2–3), 295–303. <https://doi.org/10.1016/j.schres.2005.07.018>
- Wyse, A. E. (2013). DIF Cancellation in the Rasch Model. *Journal of Applied Measurement*, 14(2), 118–128. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/23816591>
- Zeidner, M., Matthews, G., & Roberts, R. D. (2004). Emotional Intelligence in the Workplace: A Critical Review. *Applied Psychology*, 53(3), 371–399. <https://doi.org/10.1111/j.1464-0597.2004.00176.x>
- Zwick, R., Thayer, D. T., & Lewis, C. (1999). An Empirical Bayes Approach to Mantel-Haenszel DIF Analysis. *Journal of Educational Measurement*, 36(1), 1–28. <https://doi.org/10.1111/j.1745-3984.1999.tb00543.x>

# Rasch Model analysis of the Situational Test of Emotional Understanding – Brief in a large Portuguese sample

Carolina da Motta<sup>a), b)</sup>, Paula Castilho<sup>a)</sup>, Michele T. Pato<sup>c)</sup>, Célia Barreto Carvalho<sup>a), b), c)</sup>

a) Center for Research in Neuropsychology and Cognitive and Behavioral Intervention (CINEICC), College of Psychology and Education Sciences, University of Coimbra, Portugal

b) Faculty of Social and Human Sciences, University of Azores, Azores, Portugal

c) SUNY Downstate Medical College, Brooklyn, NY

## Authors' e-mails and ORCID:

Carolina da Motta – [carolina.d.motta@fpce.uc.pt](mailto:carolina.d.motta@fpce.uc.pt), <https://orcid.org/0000-0001-8421-2956>; Paula Castilho - [paulacastilho@fpce.uc.pt](mailto:paulacastilho@fpce.uc.pt), <https://orcid.org/0000-0003-1864-3146>; Michele T. Pato - [michele.pato@downstate.edu](mailto:michele.pato@downstate.edu), <https://orcid.org/0000-0002-4589-7398>; Célia Barreto Carvalho - [celia.mo.carvalho@uac.pt](mailto:celia.mo.carvalho@uac.pt), <https://orcid.org/0000-0003-4453-8139>

Correspondence concerning this article should be addressed to: Carolina da Motta, CINEICC, Faculdade de Psicologia e de Ciências da Educação da Universidade de Coimbra, Rua do Colégio Novo, Apartado 6153, 3001-802, Coimbra, Portugal. E-mail: [carolina.d.motta@fpce.uc.pt](mailto:carolina.d.motta@fpce.uc.pt) (Carolina da Motta)



## **Compliance with Ethical Standards**

**Conflict of Interest:** The authors declare that they have no conflict of interest.

## **Acknowledgments**

This research was supported by the first author's (Carolina da Motta) Ph.D. Grant (SFRH/BD/110308/2015), sponsored by FCT (Foundation for Science and Technology), Portugal, and co-sponsored by ESF (European Social Fund), Brussels, through Portuguese POPH (Human Potential Operational Program). The authors would like to thank the undergraduate students for their aid in data collection and JR for data collection and preparation.

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author upon reasonable request.

Table 1.  
*Global fit statistic of STEU-B (N = 899)*

	Reliability	Separation		Measure	Model		
				( $\theta$ )	Error	Infit	Outfit
Person fit	.68	1.45	<i>M</i>	-.01	.56	.98	1.19
			<i>SD</i>	1.00	.07	.25	1.11
			Max	3.57	1.08	1.79	9.90
			Min	-3.38	.52	.48	.26
Item fit	1.00	14.34	<i>M</i>	.00	.08	.98	1.22
			<i>SD</i>	1.19	.01	.09	.66
			Max	2.41	.11	1.14	3.61
			Min	-1.58	.07	.85	.78

*Note.* Max = maximum; Min = minimum.

Table 2.

*Global fit statistic of STEU-B (N =867)*

		$\theta$	<b>Model Error</b>	<b>Infit</b>	<b>Outfit</b>	<b>Reliability (lower- upper bound)</b>	<b>Separation (lower- upper bound)</b>
Person fit	<i>M</i>	.06	.56	.99	1.05	.60-.64	1.23-1.33
	<i>SD</i>	.94	.06	.25	.58		
	Max	3.67	1.09	1.83	4.36		
	Min	-3.42	.53	.45	.24		
Item fit	<i>M</i>	.00	.08	.99	1.05	1.00-1.00	14.51-14.74
	<i>SD</i>	1.27	.01	.08	.22		
	Max	2.69	.12	1.15	1.63		
	Min	-1.65	.07	.86	.81		

*Note.* Max = maximum; Min = minimum.

Table 3.

*Total score,  $\theta$ , Standard Error, Infit e Outfit by Item, and Point-bisserial between item and total score of STEU-B ( $n = 867$ )*

Item	Total score	$\theta$	$SE$	Infit	Outfit	$r$
1	602	-.88	.08	1.09	1.13	.391
2	261	1.06	.08	1.04	1.07	.330
3	77	2.69	.12	1.00	1.07	.205
4	88	2.53	.12	1.06	1.46	.217
5	378	.39	.07	1.15	1.33	.370
6	705	-1.65	.09	1.03	1.05	.365
7	567	-.66	.08	.86	.81	.394
8	629	-1.06	.08	.88	.83	.388
9	681	-1.44	.09	.89	.82	.375
10	617	-.98	.08	.92	.91	.390
11	429	.12	.07	.90	.88	.382
12	497	-.26	.07	.95	.92	.391
13	706	-1.65	.09	.88	.88	.365
14	439	.06	.07	1.00	.99	.383
15	575	-.71	.08	1.06	1.06	.393
16	285	.92	.08	1.08	1.28	.340
17	541	-.51	.08	.99	.97	.394
18	164	1.75	.09	1.09	1.63	.279
19	401	.27	.07	.97	.97	.376
<i>M</i>	454.8	.00	.08	.99	1.05	
<i>SD</i>	195.9	1.27	.01	.08	.22	

*Note.*  $\theta$  = measure;  $SE$  = standard error;  $r$  = point-measure correlation (between observations and  $\theta$ ).

Table 4.  
Differential item functioning between youths and adults (n = 867)

Item #	Adults				Youths				Rasch-Welch		Mantel-Haenszel	
	Obs-Exp Average	DIF Measure	DIF S.E.	Obs-Exp Average	DIF Measure	DIF S.E.	DIF Contrast	t	p	$\chi^2$	p	
1	-.030	-.710	.120	.020	-1.000	.100	.290	1.790	.074	.630	.427	
2	.020	.970	.110	-.020	1.160	.110	-.200	-1.230	.218	2.814	.093	
3	.000	2.740	.180	.000	2.640	.170	.100	.400	.687	.048	.827	
4	-.050	3.140	.210	.040	2.120	.140	1.020	4.090	.000	7.113	.008	
5	-.080	.740	.110	.060	.120	.100	.620	4.160	.000	6.931	.009	
6	-.060	-1.200	.140	.040	-1.950	.130	.750	3.990	.000	13.610	.000	
7	.040	-.900	.130	-.030	-.510	.100	-.390	-2.380	.018	.750	.387	
8	.030	-1.280	.140	-.020	-.930	.100	-.350	-1.970	.049	.454	.500	
9	.060	-2.030	.180	-.040	-1.180	.110	-.860	-4.070	.000	1.344	.001	
10	-.010	-.910	.130	.010	-1.020	.100	.110	.640	.526	2.025	.155	
11	.070	-.230	.110	-.050	.380	.100	-.610	-4.030	.000	11.666	.001	
12	.060	-.540	.120	-.040	-.050	.100	-.490	-3.160	.002	7.232	.007	
13	.060	-2.320	.200	-.040	-1.380	.110	-.940	-4.100	.000	11.084	.001	
14	.010	.010	.110	-.010	.100	.100	-.090	-.570	.566	.078	.780	
15	-.010	-.680	.120	.000	-.730	.100	.060	.350	.724	.024	.877	
16	-.090	1.390	.120	.070	.550	.100	.830	5.290	.000	16.445	.000	
17	.030	-.650	.120	-.020	-.410	.100	-.230	-1.490	.137	1.518	.218	
18	-.080	2.290	.150	.050	1.350	.110	.940	4.940	.000	11.073	.001	
19	.020	.160	.110	-.020	.350	.100	-.190	-1.280	.200	.832	.362	

Table 5.

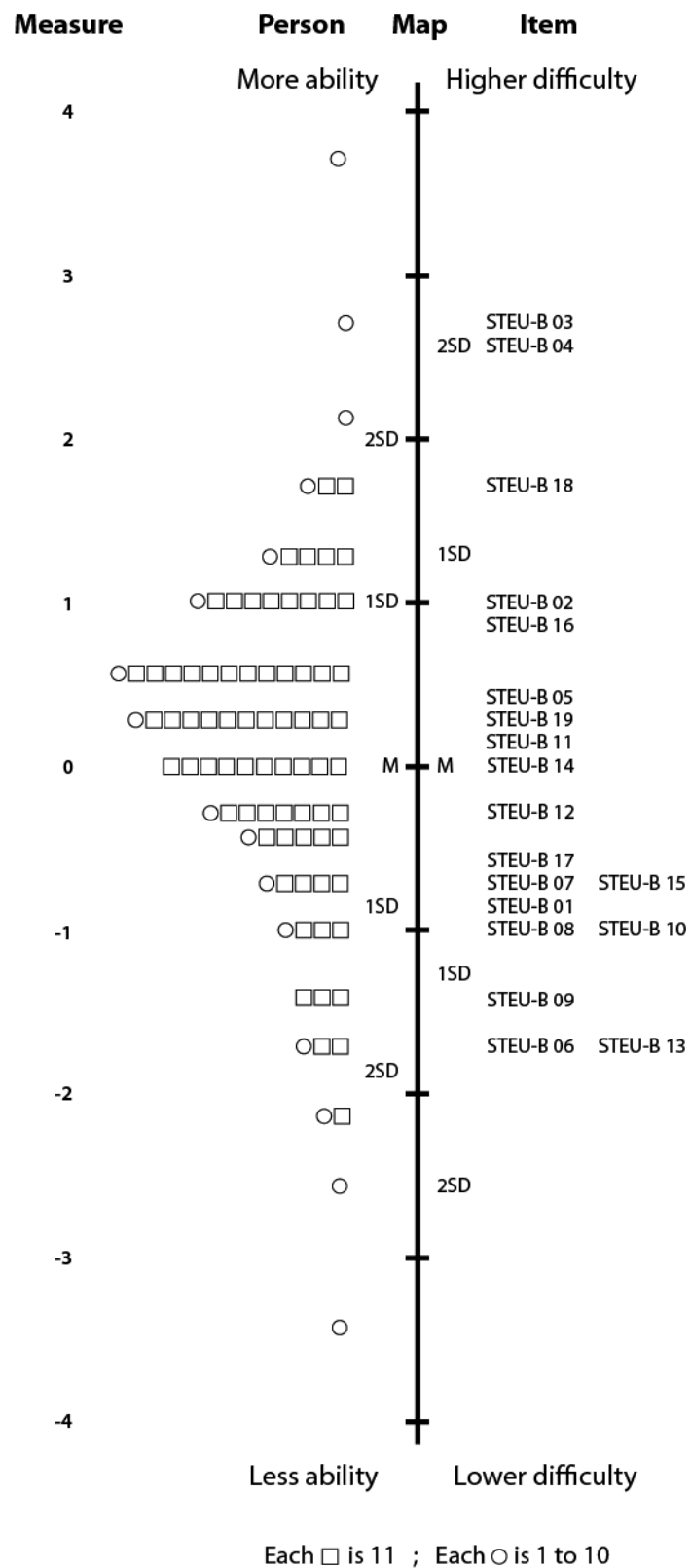
Table of sample norms (500/100) and frequencies corresponding to complete test

SCORE	θ	S.E.	MEASURE*	S.E.	INFO	FREQ	FREQ%	CUM. FREQ	CUM FREQ (%)	PERCENTILE
0	-4.7	1.85	3	19	0.29	0	0	0	0	0
1	-3.43	1.05	16	10	0.91	6	0.7	6	0.7	1
2	-2.63	0.78	24	8	1.65	5	0.6	11	1.3	1
3	-2.12	0.67	29	7	2.26	15	1.7	26	3	2
4	-1.72	0.6	33	6	2.73	32	3.7	58	6.7	5
5	-1.37	0.57	36	6	3.1	33	3.8	91	10.5	9
6	-1.07	0.55	39	5	3.36	38	4.4	129	14.9	13
7	-0.78	0.53	42	5	3.53	51	5.9	180	20.8	18
8	-0.5	0.53	45	5	3.62	57	6.6	237	27.3	24
9	-0.22	0.53	48	5	3.63	86	9.9	323	37.3	32
10	0.06	0.53	51	5	3.56	110	12.7	433	49.9	44
11	0.34	0.54	53	5	3.43	122	14.1	555	64	57
12	0.64	0.56	56	6	3.24	138	15.9	693	79.9	72
13	0.96	0.58	60	6	2.99	96	11.1	789	91	85
14	1.31	0.61	63	6	2.68	50	5.8	839	96.8	94
15	1.71	0.66	67	7	2.33	24	2.8	863	99.5	98
16	2.18	0.72	72	7	1.92	2	0.2	865	99.8	99
17	2.78	0.83	78	8	1.44	1	0.1	866	99.9	99
18	3.67	1.09	87	11	0.84	1	0.1	867	100	99
19	5.01	1.88	100	19	0.28	0	0	867	100	100

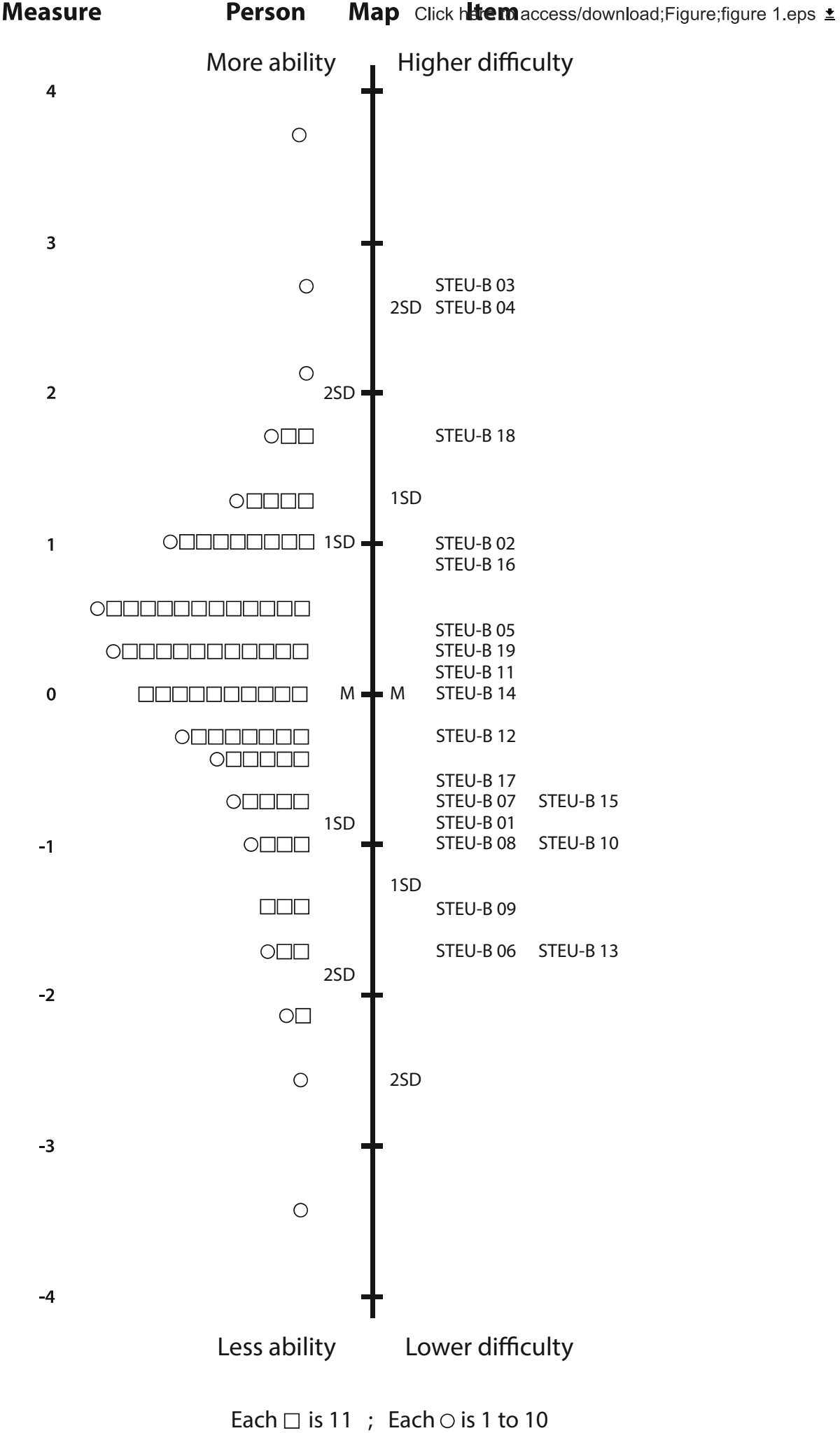
**Note:** \*θ converted to 0-100 range.

Maximum statistically different levels of performance (strata) = 3.3

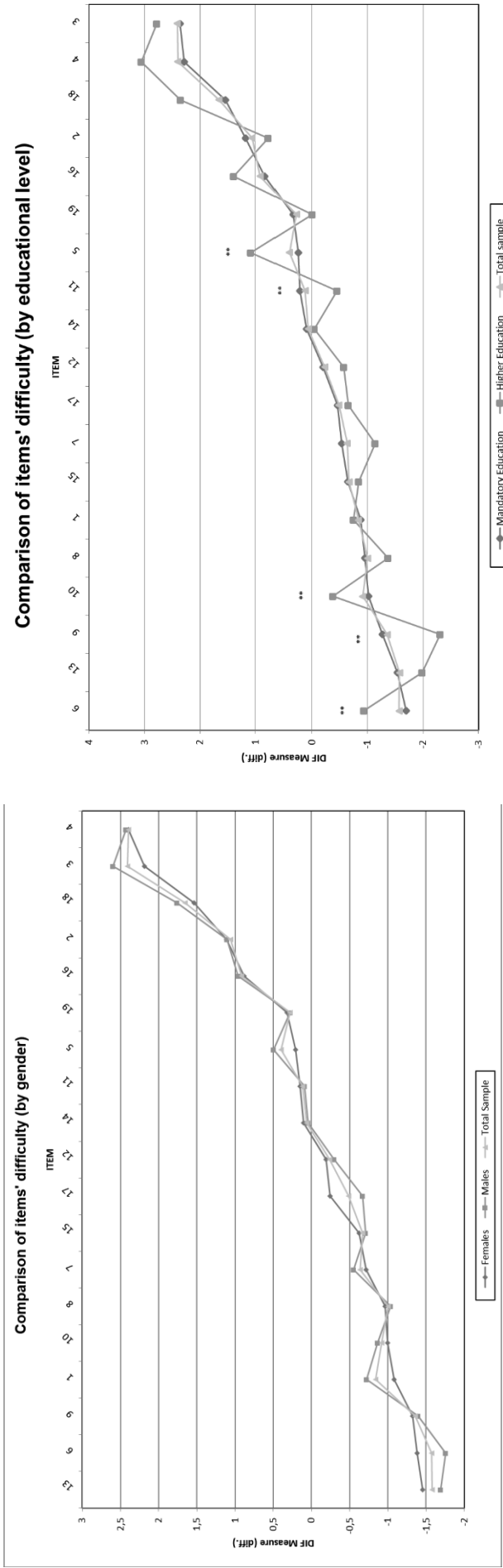
Wright's Sample-independent Person (Test) Reliability based on maximum strata = .92



**Measure** **Person** **Map** **Item** [Click here to access/download;Figure;figure 1.eps](#) 

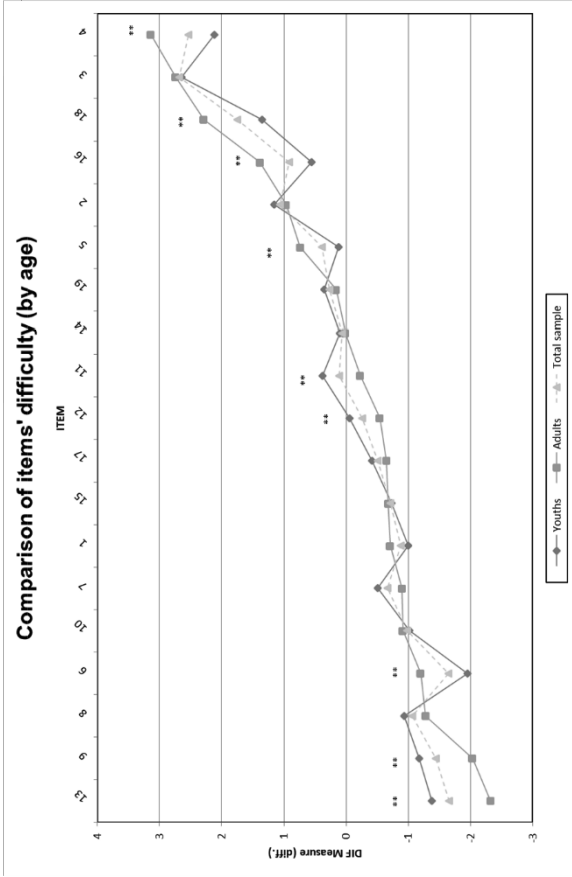






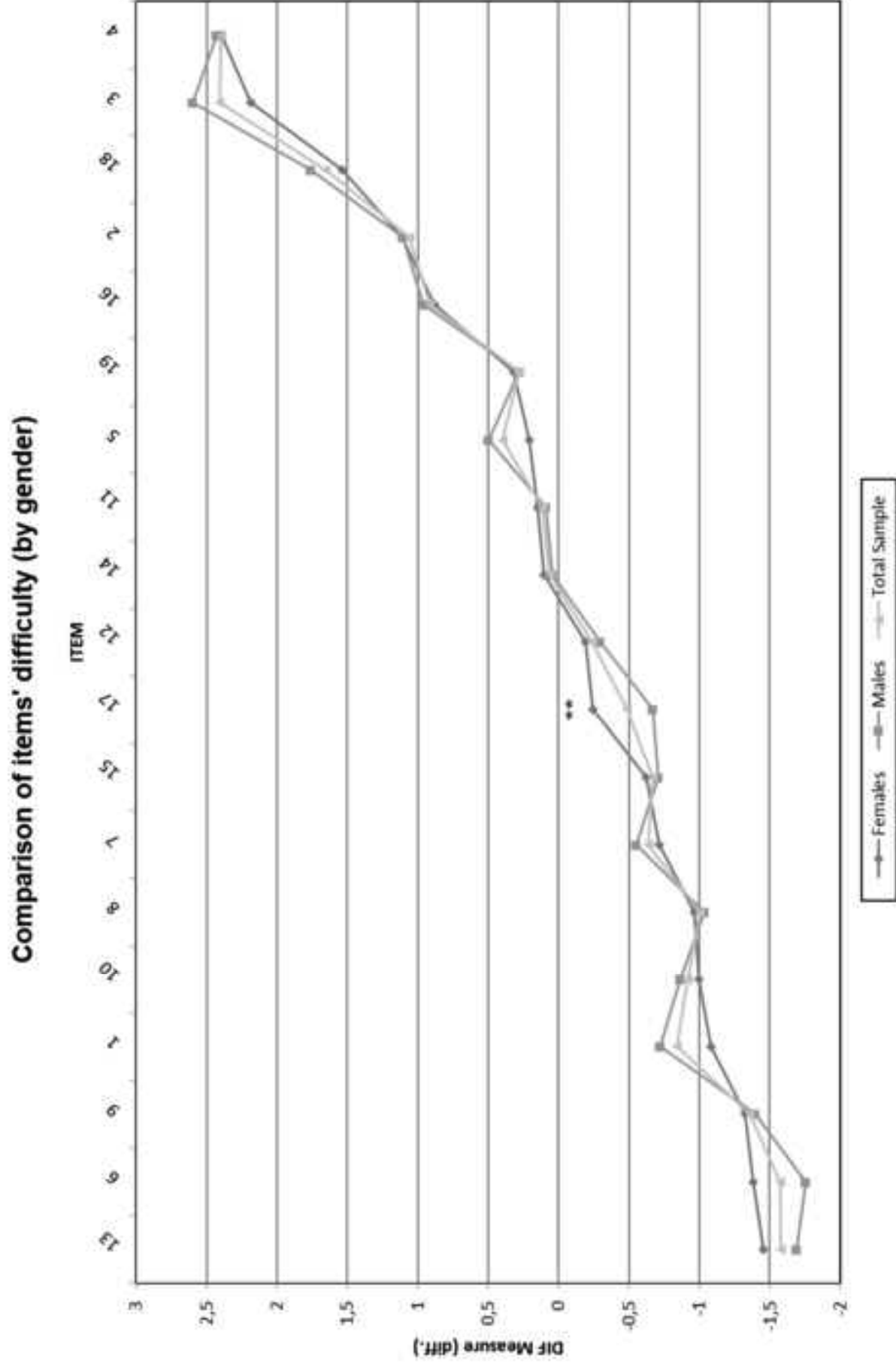
(A)

(B)

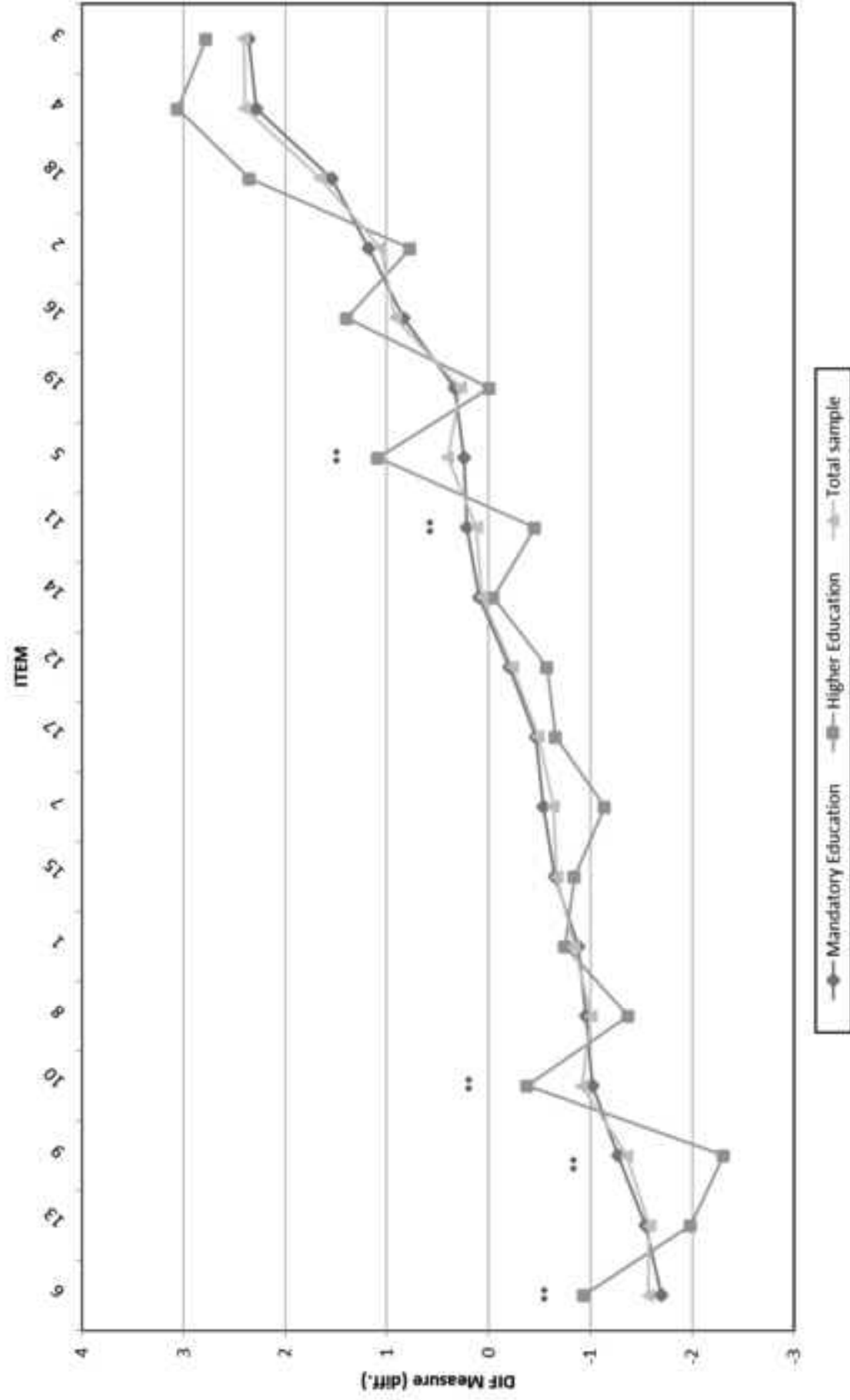


(C)

Figure 2. Changes in item difficulty between male and female participants (A), level of education (B) and age (C) (n = 867); \*\*:  $p \leq .001$

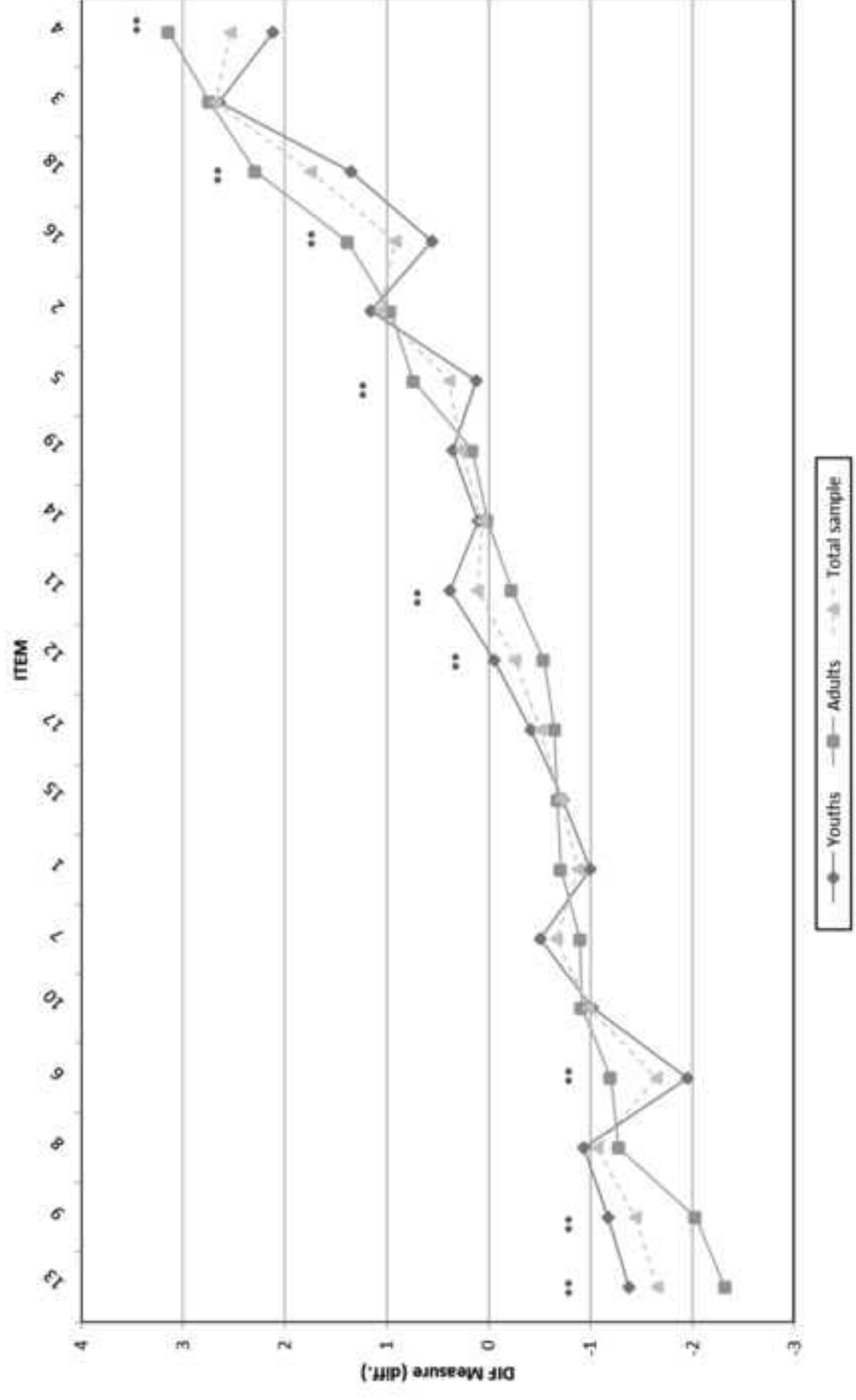


Comparison of items' difficulty (by educational level)



B

Comparison of items' difficulty (by age)



C