Emotional Self-efficacy Scale: A Categorical Confirmatory Factor Analysis with Different Numbers of Response Categories

Mohammed A. A. Abulela

University of Minnesota

April 13, 2024

Paper presented at the annual meeting of the

American Educational Research Association, Philadelphia, PA.

Citation:

Abulela, M. A. A. (2024, April 13). *Emotional self-efficacy scale: A categorical confirmatory factor analysis with different numbers of response categories* [Paper presentation]. American Educational Research Association Annual Meeting, Philadelphia, PA.

Emotional Self-efficacy Scale: A Categorical Confirmatory Factor Analysis with Different Numbers of Response Categories

Background

Since Bandura's seminal paper was published in 1977, self-efficacy theory has been widely studied in educational and psychological research (e.g., Alessandri et al., 2009). Following self-efficacy theory, domain-specific concepts have been studied. One of these concepts is emotional self-efficacy (ESE), which is an aspect of trait emotional intelligence and describes individuals' beliefs about their ability to regulate and manage self-emotions and understand those of others (Kirk et al., 2008).

Due to the increased interest in studying emotions, ESE has received much attention in the extant literature (Caprara et al., 2008; Loeb et al., 2016; Mesurado et al., 2018). Kirk and colleagues (2008) developed a scale for measuring ESE consistent with the four-component model of emotional intelligence proposed by Mayer et al. (2004). This model includes four dimensions: (a) having accurate perception of self- and others' emotions, (b) using emotions to aid thinking (i.e., decision making), (c) understanding self- and others' emotions, and (d) managing self- and others' emotions effectively. The description and validation of the score interpretations and uses of the ESE scale is discussed below.

Development of the ESE Scale

Kirk et al. (2008) generated an initial item pool following the four-aspect framework of the trait emotional intelligence illustrated earlier. Six subject matter experts reviewed the item pool and modifications were made accordingly. The final item pool comprised 32 items distributed evenly across the four aspects. To initially validate the intended score interpretations of the scale, it was administered to 207 adults (125 females, 74 males, and 8 not specified) who were recruited from various regions of Australia. For each item, participants rated their confidence on a five-point rating scale ranging from "*not at all*" to "*very*".

Using principal component analysis (PCA), Kirk and colleagues found five components with eigenvalues greater than one where the respective eigenvalues were 13.96, 1.65, 1.50, 1.21 and 1.07. Relatedly, Cattell's (1966) scree plot, the minimum average partial correlation (MAP; Velicer 1976), and parallel analysis (PA, Horn 1965) indicated one, three, and four components be retained, respectively. However, the large difference between the first and second eigenvalues as well as the recommendations of Cattell's rule suggested that one component be retained. The authors therefore adopted a one-component solution that explained 44% of the variance in item scores and consequently ESE was concluded to be a unidimensional construct.

Kirk and colleagues also collected validity evidence based on associations with other variables and found that high ESE was associated with greater dispositional emotional intelligence, greater performance emotional intelligence, higher positive mood and lower negative mood. Score reliability estimates were .96 and .85 using alpha and test-retest reliability, respectively. However, to utilize the ESE scale in other contexts, appropriate sources of validity evidence should be collected (American Educational Research Association et al., 2014), which call for more psychometric studies.

Psychometric Studies on the ESE Scale

In a subsequent validation study, Dacre Pool and Qualter (2012a) administered the 32-item ESE scale to undergraduate students (543 calibration sample, 542 validation sample). Based on the results of exploratory factor analysis (EFA) with principal axis factoring, four factors were retained that explained 54.18% of the common variance. Five items had loadings below the adopted threshold (i.e., .45) and consequently were removed from the scale. Using the validation sample, CFA was conducted on the remaining 27 items to test for the unidimensional factor structure found in previous research (Kirk et al., 2008) and the four-factor model obtained from EFA. Based on fit indices, the four-factor correlated model had better model-data fit and consequently was retained. Thus, based on the results of both EFA and CFA, ESE was found to be a multidimensional construct.

To further compare the unidimensional vs. multidimensional factor structure of the 27-item ESE scale among youth, Qualter and colleagues (2015) administered it to 192 adolescents aged 11 to 13 years from the U.K. CFA Results supported the four- factor model compared to the unidimensional model. It seems obvious that the ESE scale was developed and evaluated in western countries (e.g., Australia and the U.K.).

Limitations of Prior Psychometric Studies

Though ESE scale has been widely used in cross-sectional (e.g., Dacre Pool & Qualter, 2013) and intervention (e.g., Dacre Pool & Qualter, 2012b) studies, there have been some methodological issues associated with its psychometric validation. First, a major methodological flaw has been using PCA to explore its initial factor structure despite the well-illustrated differences between PCA and factor analysis (Fabrigar et al., 1999). Relatedly, since PCA uses "ones" in the diagonal of the correlation matrix to account for both common and unique variance; loading, communalities, and amount of variance explained are overestimated and consequently relatively wrong conclusions (Park et al., 2002). Second, to test the one- versus the four factor model, prior researchers utilized confirmatory factor analysis (CFA) with maximum likelihood for ordinal data, which did not possess multivariate normality and consequently results might be biased (e.g., Dacre Pool & Qualter, 2012a). Third, the extent to which fit indices and magnitude of standardized loadings might differ, as a function of the different numbers of response categories (NRC) has not been well investigated. The last is of a particular interest, given the potential impact of NRC on CFA results, which is discussed next.

The NRC and CFA Results

The choice of the appropriate NRC when developing and validating score interpretations of new measures has received considerable attention (Asu´n et al., 2016; Barnette, 2010; Danner et al., 2016). One of the primary controversial issues has been the inclusion of the midpoint option, where some researchers have supported its use (e.g., Hurley, 1998), others have discouraged so due to yielding inaccurate responses (Garland, 1991). Some researchers have taken a middle position indicating the inclusion or exclusion of the midpoint should be carefully considered taking into account the phenomena being studied (e.g., Nadler et al., 2015).

Before diving into the specific impact of the NRC on CFA results, it is important to discuss how range restriction generally affects factor analysis outputs. In a recent simulation study, Franco-Martı´nez and colleagues (2023) manipulated sample size (200 and 500 cases), test length (6, 12, 18, and 24 items), magnitude of loading (.50, .70, and .90), and restriction size (from R = 1, .90, .80, and so on till .10 selection ratio). The authors found that loading were underestimated as range restriction increased. Range restriction might be attributed to fewer NRC (e.g., three) due to the less variability in responses. With regard to CFA, previous researchers have found an impact for NRC on CFA fit indices. For instance, Dolan (1994) found that chi-square value tended to decrease as the NRC increases. In contrast, Maydeu-Olivares et al. (2009) found that the smaller the NRC, the better model-data fit. For a comprehensive review about the impact of NRC on validity evidence, interested readers are referred to Abulela and Khalaf (2024).

Current Study

The current study contributes to this ongoing research endeavors by addressing the gaps and methodological limitations noted earlier. Specifically, we addressed such limitations by utilizing categorical CFA with the weighted least squares mean and variance adjusted (WLSMV) estimator as a robust methodology for ordinal data (Beauducel & Herzberg, 2006; Wirth & Edwards, 2007; Xing & Hall, 2015). In brief, three research questions are under investigation:

1. Is the ESE scale unidimensional or multidimensional (e.g., four correlated factors) within Egyptian undergraduate students?
2. Which NRC (e.g., three, five, and seven) results in better model data fit?
3. Does the number of items with low standardized loadings (e.g., < .30) differ according to NRC and the factor structure fitted?

Methods

Participants

A total of 243 undergraduate students (Age range: 20-22 years, $M_{age} = 20.22$, $SD_{age} = .44$), enrolled in a large public university in Upper Egypt, participated voluntarily in the study. We followed all ethical guidelines of research on human participants during the scale administration.

Measures

As stated earlier, the ESE scale was originally developed in Australia. To translate the scale into Arabic and adapt it for use in the present study, we utilized the back-translation technique (Brislin, 1970). To ensure the proper implementation of this technique, we followed some steps. First, two bilingual professors were invited to independently translate the scale into Arabic. Then, we met and compared the two translated versions, where minor discrepancies were discussed and resolved. Next, a third bilingual professor back translated the Arabic version into English. Last, we rigorously compared the original with the back translated version of the scale. After comparison, we found that the two versions were highly comparable, indicating accuracy of translation.

Procedures

To examine the impact of NRC on the results of categorical CFA, three forms of the ESE scale were created differing only in the NRC (three, five, and seven). To ensure different results are due to varying the NRC rather than a particular cohort is responding to each form of the scale, a repeated measures design was utilized to account for intra- individual effects. Following this design, each participant completed the three Arabic forms of the ESE scale in one session with a five-minute break before each form. Participants rated each statement on a three, five, and seven point rating scale.

Data Analysis

Applied researchers and psychometricians are advised to fit rival factor models when collecting validity evidence-based on the internal structure of the measure (American Educational Research Association et al., 2014). This assists in selecting the final model consistent with the conceptual framework, practical evidence, and the intended interpretations and uses of scores (i.e., unidimensional vs. multidimensional factor structure).

Ceiling and Floor Effects

Prior to conducting data analyses, it is recommended to examine response frequencies per item to check for ceiling and floor effects (Bandalos, 2018). Ceiling and floor effects are present when more than 15% of participants select the highest or lowest response option, respectively (McHorney & Tarlov, 1995).

Categorical CFA

To answer the first research question and assess the dimensionality of ESE scale, two competing models were fitted using categorical CFA appropriate for ordinal data with WLSMV estimator available in Mplus (Muthén & Muthén, 1998-2017). We fitted one- and four-factor categorical CFA models to establish validity evidence-based on internal structure of the ESE in a non-western population.

To evaluate the fit of the tested models, we utilized Marsh's et al. (2004) and others' guidelines since the restrictive indices proposed by Hu and Bentler (1999) had some limitations, since they were originally proposed for continuous data. Overall, models with root mean square error of approximation (RMSEA) < 0.05 have good fit, values between .05 and .08 indicate reasonable fit, values between .08 and 1 suggest poor fit, and values > .1 are unacceptable (Qualter et al., 2015). Relatedly, comparative fit index (CFI) and Tucker-Lewis index (TLI) ≥ .90 are indicative of good fit (Hoe, 2008). Standardized root mean squared residual (SRMR) ≤ .08 suggests close fit (Asparouhov & Muthén, 2018). When two competing nested models have good fit, the chi-square difference test ($\chi^2_{Diff.}$), change in CFI (ΔCFI), and change in RMSEA (ΔRMSEA) are utilized to select the model with the best fit.

Results

Ceiling and Floor Effects

After computing frequencies, no substantial evidence for ceiling or floor effects was observed, since most item response frequencies did not exceed 15%.

Research Question 1: Assessing Dimensionality of the ESE Scale

Despite the lack of fit for the two rival unidimensional and four-factor correlated models, fit indices were in favor of the multidimensional structure of the ESE scale, particularly RMSEA values that were within the acceptable range. Table 1 provides fit indices for the two competing models. Given that none of the two competing models had adequate model data fit, we did not conduct model comparison.

Table 1

Fit Indices of the One-and Four Correlated-factor Model of Emotional Self-efficacy According to Three, Five, and Seven Response Categories

Model	NRC	χ^2	CFI	TLI	SRMR	RMSEA	RMSEA 90 % CI
One-factor	Three	1195.93**	0.342	0.287	0.15	0.105	0.099 – 0.112
	Five	1725.30**	0.406	0.357	0.151	0.134	0.128 – 0.140
	Seven	1740.09**	0.443	0.396	0.142	0.134	0.128 – 0.141
Four-correlated Factors	Three	716.638**	0.699	0.668	0.122	0.072	0.065 – 0.079
	Five	884.713**	0.760	0.735	0.112	0.086	0.079 – 0.093
	Seven	889.539**	0.775	0.752	0.104	0.086	0.08 – 0.093

Note. NRC = number of response categories, χ^2 = chi-square statistics, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, SRMR = Standardized Root Mean Square Residual, RMSEA = Root Mean Square Error of Approximation, CI = RMSEA 90 %confidence intervals. Degrees of freedom were 324 and 318 for the one- and four-factor model, respectively.

**P < .001

Research Question 2: The NRC and Fit Indices

As shown in Table 1, fit indices behaved differently as the NRC increased. Specifically, the chi-square and RMSEA tended to increase indicating worse fit. Conversely, CFI and TLI values increased but SRMR decreased as the NRC increased, suggesting a relatively better fit. This pattern was observed across the unidimensional and multidimensional factor structures. These results call for not relying solely on fit indices to accept and reject hypothesized models as recommended in previous literature. In particular, standardized loadings, factor correlations, and residual variance should be examined (Brown, 2006).

Research Question 3: The NRC and Number of Items with Low Standardized Loadings

As noted earlier, standardized loadings should be carefully examined when conducting CFA analyses. Overall, the number of items that had standardized loadings below .30 decreased as the NRC increased. In case of the unidimensional model, the respective number of items with low loadings was 13, 10, and eight items for the three, five, and seven response categories. However, with the multidimensional model, the number of items with low loadings decreased substantially regardless of the NRC. Specifically, this number became two, two, and one for the three, five, and seven response categories, respectively.

Table 2

Standardized Factor Loadings of the One- and Four Correlated-factor Categorical CFA Models According to Three, Five, and Seven Response Categories

Item No.	One-factor Model			Four Correlated-factor Model		
NRC	Three	Five	Seven	Three	Five	Seven
8	0.347**	**0.215**	**0.222**	0.443	0.355	0.376
3	**0.255****	**0.23**	**0.255**	0.36	0.351	0.413
18	0.32**	0.354	0.423	0.57	0.667	0.729
30	**0.294**	**0.188**	**0.241**	0.391	0.37	0.429
14	**0.147***	**0.148**	**0.223**	<u>**0.155**</u>	**0.283**	0.305
12	**0.242**	<u>**0.058**</u>	**0.157**	0.421	0.449	0.425
22	0.383	0.34	0.35	0.57	0.594	0.577
20	**0.235**	<u>**0.034**</u>	**0.113***	0.446	0.469	0.424
6	0.395	0.394	0.418	0.597	0.598	0.589
26	**0.297**	<u>**0.26**</u>	0.313	0.422	0.442	0.471
27	**0.142***	0.648	0.644	0.716	0.801	0.763
11	<u>**0.088**</u>	0.661	0.629	0.695	0.811	0.78
1	**0.156***	0.508	0.591	0.598	0.642	0.714
9	0.397	0.571	0.601	0.707	0.665	0.703
19	**0.291**	0.543	0.599	0.559	0.672	0.692
4	0.354	0.539	0.517	0.665	0.655	0.624
7	0.554	0.547	0.514	0.633	0.666	0.661
24	0.39	**0.212**	0.324	0.432	0.423	0.521
31	0.623	0.543	0.475	0.712	0.713	0.634
32	0.559	0.371	0.449	0.579	0.6	0.637
15	0.649	0.565	0.43	0.735	0.69	0.596
2	0.487	0.35	0.443	0.504	0.575	0.613
23	0.596	0.525	0.492	0.656	0.621	0.588
13	0.555	0.554	0.472	0.6	0.648	0.614
25	**0.269**	0.338	0.33	0.741	0.797	0.846
21	<u>**0.074**</u>	<u>**0.106**</u>	**0.193**	<u>**0.195**</u>	**0.203***	**0.294**
17	**0.2**	**0.296**	**0.237**	0.722	0.683	0.642
No. of items with low Standardized loadings	13	10	8	2	2	1

Note. NRC = number of response categories, Bold loadings define those below 0.30. Bold and underlined loadings are nonsignificant.

Given that the multidimensional model fits the data relatively better, this suggests that the impact of the NRC on the number of low standardized loadings is moderated by the misspecification of the factor structure. Stated differently, as the degree of model misspecification increases, the impact of the NRC on the number of items with low loadings increases. Relatedly, the magnitude of standardized loadings tends to be higher with the less misspecified factor structure and more response categories. This is clearly reflected in the multidimensional model with seven response categories. This indicates that more interpretable factors explain more variations in item responses.

Discussion

ESE has gained much interest in educational research due to the role of emotions in learning and instruction. The overall objective of the present study was to assess the dimensionality of the ESE scale and investigate if fit indices and number of items with low standardized loadings differ based on utilizing three, five, and seven response options. The first objective was pursued to collect validity evidence based on the internal structure for the ESE scale in Egyptian undergraduates, utilizing the WLSMV estimator as appropriate for ordinal data needed to obtain unbiased parameter estimates. Based on this analysis, we emphasize that applied researchers and psychometricians should utilize psychometric models (e.g., categorical CFA) appropriate for the type of data being analyzed. Based on the study results, neither the one- nor the four-factor model fits the data well, with the multidimensional factor structure having relatively better fit.

Regarding the effect of the NRC on fit indices as well as the number of items with low standardized loadings, we found important results since the performance of fit indices was not uniform, meaning that different types of fit indices behaved differently as the NRC increased. Most importantly, the number of items with low loadings was more impacted by the misspecification of the factor structure compared to the NRC. This cautions against the removal of items with low loadings with further discussion with subject matter experts, which might threaten validity evidence based on test content (i.e., construct underrepresentation). In more detail, a researcher may remove an item because of low loading, which in fact is a result of a misspecified factor structure or a small NRC rather than being irrelevant to the construct being assessed. This highlights the role of model misspecification and NRC should be carefully considered when conducting CFA analyses.

To illustrate, fewer response categories reduce variability in item responses, which leads to correlation attenuation, which in turn yields small loadings. In addition to the misspecification of the factor structure, as the NRC decreased, the number of items with low standardized loadings increased. To conclude, the NRC should be selected carefully as being important for decision making in CFA results. Relatedly, the scale factor structure should be specified based on prior theory and score interpretation and uses.

References

Abulela, M. A. A., & Khalaf, M. A. (2024). Does the number of response categories impact validity evidence in self-report measures? A scoping review. *SAGE Open, 14*(1), 1-16. https://doi.org/10.1177/21582440241230363

Alessandri, G., Caprara, G.V., Eisenberg, N., & Steca, P. (2009). Reciprocal relations among self-efficacy beliefs and prosociality across time. *Journal of Personality, 77*(4), 1229-1259. https://doi.org/10.1111/j.1467-6494.2009.00580.x

Alessandri, G., Vecchione, M., & Caprara, G. V. (2015). Assessment of regulatory emotional self-efficacy beliefs: A review of the status of the art and some suggestions to move the field forward. *Journal of Psychoeducational Assessment, 33*(1), 24-32. https://doi.org/10.1177/0734282914550382

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing* (6th ed.). American Educational Research Association.

Asparouhov, T., & Muthén, B. (2018). SRMR in Mplus. Retrieved from https://www.statmodel.com/download/SRMR2.pdf

Asu´n, R. A., Rdz-Navarro, K., & Alvarado, J. M. (2016). Developing multidimensional Likert scales using item factor analysis: The case of four-point items. *Sociological Methods & Research*, *45*(1), 109-133. https://doi.org/10.1177/0049124114566716

Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. The Guilford Press.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191-215. https://doi.org/10.1037/0033-295X.84.2.191

Barnette, B. A. (2010). Likert scaling. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 715-718). Sage Publication Inc.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13*(2), 186-203. https://doi.org/ 10.1207/s15328007sem1302_2

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*(3), 185-216. https://doi.org/10.1177/135910457000100301

Brown, T. A. (2006). *Confirmatory factor analysis for applied research* (1st ed.). The Guilford Press.

Caprara, G. V., Di Giunta, L., Eisenberg, N., Gerbino, M., Pastorelli, C., & Tramontano, C. (2008). Assessing regulatory emotional self-efficacy in three countries. *Psychological Assessment, 20*(3), 227-237. https://doi.org/ 10.1037/1040-3590.20.3.227

Dacre Pool, L., & Qualter, P. (2012a). The dimensional structure of the emotional self-efficacy scale (ESES). *Australian Journal of Psychology, 64*, 147-154. https://doi.org/ 10.1111/j.1742-9536.2011.00039.x

Dacre Pool, L., & Qualter, P. (2012b). Improving emotional intelligence and emotional self-efficacy through a teaching intervention for university students. *Learning and Individual Differences, 22*(3), 306-312. https://doi.org/10.1016/j.lindif.2012.01.010

Dacre Pool, L., & Qualter, P. (2013). Emotional self-efficacy, graduate employability, and career satisfaction: Testing the associations. *Australian Journal of Psychology, 65*, 214-223. https://doi.org/10.1111/ajpy.12023

Danner, D., Blasius, J., Breyer, B., Eifler, S., Menold, N., Paulhus, D. L., Rammstedt, B., Roberts, R. D., Schmitt, M., & Ziegler, M. (2016). Current challenges, new developments, and future directions in scale construction [Editorial]. *European Journal of Psychological Assessment, 32*(3), 175–180. https://doi.org/10.1027/1015-5759/a000375

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology, 41*(2), 309-326. https://doi.org/10.1111/j.2044-8317.1994.tb01039.x

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Franco-Martínez, A., Alvarado, J. M., & Sorrel, M. A. (2023). Range restriction affects factor analysis: Normality, estimation, fit, loadings, and reliability. *Educational and Psychological Measurement, 83*(2), 262–293. https://doi.org/10.1177/00131644221081867

Galla, B. M., &Wood, J. J. (2012). Emotional self-efficacy moderates anxiety-related impairments in math performance in elementary school-age youth. *Personality and Individual Differences, 52*(2), 118–122. https://doi.org/10.1016/j.paid.2011.09.012

Garland, R. (1991). The mid-point on a rating scale: Is it desirable? *Marketing Bulletin, 2*, 66-70.

Goroshit, M., & Hen, M. (2014). Does emotional self-efficacy predict teachers' self-efficacy and empathy? *Journal of Education and Training Studies, 2*(3), 26-32. http://dx.doi.org/ 10.11114/jets.v2i 3 3 59

Green, S. B., Akey, T. M., Fleming, K. K., Hershberger, S. L., & Marquis, J. G. (1997). Effect of the number of scale points on chi-square fit indices in confirmatory factor analysis. *Structural Equation Modeling, 4*(2), 108-120. https://doi.org/10.1080/10705519709540064

Hall, A. J. (2017). *Dimensionality and instrument validation in factor analysis: Effect of the number of response alternatives* (Master’s thesis). Available from ProQuest Dissertations and Theses database. (UMI No. 10264211)

Hoe, S. L. (2008). İssues and procedures in adopting structural equation modelling technique. *Journal of Applied Quantitative Methods, 3*(1), 76-83.

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

Hurley, J. R. (1998). Timidity as a response style to psychological questionnaires. *The Journal of Psychology: Interdisciplinary and Applied, 132*(2), 201-210. https://doi.org/10.1080/00223989809599159

Kirk, B. A., Schutte, N. S., & Hine, D. W. (2008). Development and preliminary validation of an emotional self-efficacy scale. *Personality and Individual Differences, 45*(5), 432-436. https://doi.org/10.1016/j.paid.2008.06.010

Loeb, C., Stempel, C., & Isaksson, K. (2016). Social and emotional self-efficacy at work. *Scandinavian Journal of Psychology, 57*, 152–161. https://doi.org/10.1111/sjop.12274

Marsh, H. W., Hau, K. T., & Wen, Z. L. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralising Hu & Bentler (1999) findings. *Structural Equation Modeling, 11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2

Maydeu-Olivares, A., Kramp, U., García-Forero, C., Gallardo-Pujol, D., & Coffman, D. (2009). The effect of varying the number of response alternatives in rating scales: Experimental evidence from intra-individual effects. *Behavior Research Methods, 41*(2), 295-308. https://doi.org/10.3758/BRM.41.2.295

Maydeu-Olivares, A., Fairchild, A. J. & Hall, A. G. (2017). Goodness of fit in item factor analysis: Effect of the number of response alternatives. *Structural Equation Modeling, 24*(4), 495-505. https://doi.org/10.1080/10705511.2017.1289816

Mayer, J. D., Salovey, P., & Caruso, D. R. (2004). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry, 15*(3), 197–215. https://doi.org/10.1207/s15327965pli1503_02

McHorney, C. A., & Tarlov, A. R. (1995). Individual-patient monitoring in clinical practice: Are available health status surveys adequate? *Quality of Life Research*, *4*(4), 293–307. https://doi.org/10.1007/BF01593882

Mesurado, B., Vidal, E. M., & Mestre, A. L. (2018). Negative emotions and behaviour: The role of regulatory emotional self-efficacy. *Journal of Adolescence*, *64*, 62–71. https://doi.org/10.1016/j.adolescence.2018.01.007

Muthén, L. K., & Muthén, B. O. (1998-2017). Mplus user’s guide (8th ed.). Muthén & Muthén.

Nadler, J. T., Weston, R., & Voyles, E. C. (2015). Stuck in the middle: The use and interpretation of mid-points in items on questionnaires. *The Journal of General Psychology, 142*(2), 71-89. https://doi.org/10.1080/00221309.2014.994590

Park, H. S., Dailey, R., & Lemus, D. (2002). The use of exploratory factor analysis and principal components analysis in communication research. *Human Communication Research, 28*(4), 562–577. https://doi.org/10.1111/j.1468-2958.2002.tb00824.x

Qualter, P., Dacre Pool, L., Gardner, K. J., Ashley-Kot, S., Wise, A., & Wols, A. (2015). The Emotional self-efficacy scale: Adaptation and validation for young adolescents. *Journal of Psychoeducational Assessment*, *33*(1), 33–45. https://doi.org/10.1177/0734282914550383

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods, 12*(1), 58–79. https://doi.org/10.1037/1082-989X.12.1.58

Xing, C., & Hall, J. A. (2015). Confirmatory factor analysis and measurement invariance testing with ordinal data: An application in revising the flirting styles inventory. *Communication Methods and Measures*, *9*(3), 123-151. https://doi.org/10.1080/19312458.2015.1061651