BARÚ Bartın University Journal of Faculty of Education BUEFAD
2023, Volume 12, Issue 3, 460-470
dergipark.org.tr/buefad
DOI: 10.14686/buefad.1280435

Alternative Confirmatory Factor Analytic Models for Examining Preservice Teachers' Non-Cognitive Skills

Nilüfer Kahraman [a], Esra Sözer Boz [b*], Derya Akbaş [c], Ergün Cihat Çorbacı [d], Şerife Işık [e], Nazife Üzbe Atalay [f], Fatma Nur Aydın [g], Mehtap Çakan [h] & Şeref Sağıroğlu [*]

Research Article
Received: 10.4.2023
Revised: 24.4.2023
Accepted: 7.5.2023

a Prof. Dr., Gazi University, Türkiye, https://orcid.org/0000-0003-2523-0155
b Asst. Prof. Dr., Bartin University, Türkiye, https://orcid.org/0000-0002-4672-5264, *esrsozer@gmail.com
c Asst. Prof. Dr., Aydın Adnan Menderes University, Türkiye, https://orcid.org/0000-0001-9852-4782
d Res. Asst. Dr., Gazi University, Türkiye, https://orcid.org/0000-0002-7874-956X
e Prof. Dr., Gazi University, Türkiye, https://orcid.org/0000-0002-5184-8218
f Asst. Prof. Dr., Başkent University, Türkiye, https://orcid.org/0000-0002-6637-5855
g Mathematic Teacher, Akyurt TOKİ School, Türkiye, https://orcid.org/0000-0003-0887-395X
h Prof. Dr., Gazi University, Türkiye, https://orcid.org/0000-0001-6602-6180
i Prof. Dr., Gazi University, Türkiye, https://orcid.org/0000-0003-0805-5818

Abstract

This study provides a comparative analysis of the inferences that can be reached about preservice teachers' emotional awareness and personality traits when several alternative factor analytic models are tested simultaneously. An empirical illustration is provided using two datasets collected for a research study aiming to profile the social-emotional behavior repertoire of preservice teachers. Four alternative factor analytic models were considered for both datasets: the unidimensional model, the correlated model, the higher- (or second-) order model, and the bifactor model. Results indicate that the higher-order factor model for emotional awareness (n = 670) and the bifactor model for personality traits (n = 670) were the preferred models that better represent the underlying factor structures. Multidimensionality and practical decisions concerning score reporting are discussed.

Keywords: non-cognitive skills, teacher training, factor analytic models, multidimensionality

Öğretmen Adaylarının Bilişsel Olmayan Becerilerinin İncelenmesinde Alternatif Faktör Analitik Modeller

Öz

Bu çalışma, alternatif faktör analitik yaklaşımlarının eş zamanlı olarak test edilmesiyle öğretmen adaylarının duygusal farkındalıkları ve kişilik özellikleri hakkında yapılacak çıkarımların karşılaştırmalı bir analizini sunmaktadır. Öğretmen adaylarının sosyal-duygusal davranışlarına ilişkin profillerini belirlemeyi amaçlayan daha geniş kapsamlı bir araştırmada elde edilen iki farklı veri seti kullanılarak ampirik bir örnek sunulmuştur. Her iki veri seti için dört farklı faktör analitik model dikkate alınmıştır. Bu modeller; tek boyutlu model, ilişkili model, yüksek (veya ikinci) dereceli model ve iki faktör modelidir. Sonuçlar, duygusal farkındalık için ikinci dereceden faktör modelinin (n = 670), kişilik özellikleri için iki faktör modelinin (n = 670) veriyi daha iyi temsil eden modeller olduğunu göstermektedir. Ölçek puanlarını raporlama ve çok boyutluluk ile ilgili pratik konular tartışılmıştır.

Anahtar kelimeler: bilişsel olmayan beceriler, öğretmen eğitimi, faktör analitik modeller, çok boyutluluk.

To cite this article in APA Style:
Kahraman, N., Sözer-Boz, E., Akbaş, D., Çorbacı, E. C., Işık, Ş., Üzbe Atalay, N., Aydın, F. N., Çakan, M., & Sağıroğlu, Ş. (2023). Alternative confirmatory factor analytic models for examining preservice teachers' non-cognitive skills. *Bartın University Journal of Faculty of Education, 12*(3), 460-470. https://doi.org/10.14686/buefad.1280435

INTRODUCTION

The concurrently measuring uni-and multidimensional constructs can generate doubt about the precise nature of the measured construct(s) and bring together how researchers can report scale scores (Dunn & McCary, 2020). Therefore, examining the factor structure of tests or scales is essential for making proper validity arguments for test or scale scores. When reporting a test score on a single scale, the implication is that the test measures one unitary skill or trait and that the scores given reflect the examinee's ability or level on that single trait. Dividing the scale into sub-scales and reporting separate sub-scores indicates that each sub-score should require a sufficiently distinct aspect of ability from the other sub-scales (Dunn & McCary, 2020).

Treating multidimensional data like unidimensional data causes model misspecification, leading to biased parameter estimates that produce biased estimates of the relationships among latent variables. This can threaten the determination of actual relationships among constructs represented as latent variables. If the item response data are multidimensional, sub-scales are not interchangeable indicators of a single construct and might relate differently to an external variable. Total scale scores might not reflect the target construct of interest because multiple systematic sources of variance could confound their interpretation. In such cases, a researcher might need to consider specifying a multidimensional measurement model. The findings about multidimensionality have significant consequences in scale scoring and interpretation (Reise et al., 2013).

Psychometric analyses of measurement tools provide validity evidence that determines how precisely total and sub-scale scores reflect their underlying constructs. Another point is to determine how well a particular set of items indicates a latent variable and how the items can be used in the measurement model specification (Rodriquez et al., 2016). It is essential to decide which measurement model should be utilized when the item response data are multidimensional. The Structural Equation Modeling (SEM) framework is an appropriate analytic tool for researchers to evaluate the interrelations among a network of constructs (Reise et al., 2013). SEM assumes that the latent variable indicators reflect a single common latent variable or that all multidimensionality can be specified in more complicated measurement models, such as higher-order and bifactor models (Chen et al., 2006).

Eid and Diener (2006) suggest using multimethod approaches that offer insights into psychological structures in education and psychology. They state that there are at least two reasons for using multimethod approaches: (1) the multi-component structure of psychological phenomena and (2) the validity of the research. Psychological phenomena usually comprise many aspects. A method may be appropriate to measure one aspect but not appropriate for another aspect. To correctly measure these phenomena, it is necessary to have various research methods. Multimethod research provides information for uncovering general associations between different components and levels of phenomena. Another significant point is that multimethod research is crucial in validation (Eid & Diener, 2006). Especially concerning measurement methods, validity represents the degree to which the adequacy and appropriateness of inferences based on the results of a measurement tool are supported by empirical evidence and theoretical rationales (Messick, 1995). The multimethod approach is needed to analyze the validity appropriately, and different methods should converge in measuring the same trait, which indicates convergent validity (Campbell & Fiske, 1959). Campbell and Fiske clarify that a score on a psychological variable reflects the psychological construct under consideration and systematic method-specific influences. They demonstrate the necessity of including at least two methods in psychological studies to provide evidence for validity. Method effects can indicate valid and valuable information about the construct. Multimethod research provides proof of convergent validity and allows the analysis of the nature of method-specific influences.

Psychological phenomena such as social-emotional competencies usually comprise many aspects. In the educational system, teachers' social-emotional competence is a fundamental skill for social development and forming healthy relationships with students and others. Substantial evidence shows how teachers interact with children in a classroom affects their social and emotional outcomes (Ulloa et al., 2016). The social-emotional competence level of teachers plays a crucial role in developing positive relationships with children and contributes to the formation of healthy atmospheres in classrooms. Improving teachers' social-emotional competence reduces disruptive classroom behaviors while increasing self-control and academic achievement (Jennings & Greenberg, 2009; McCarthy, 2021). The measurement tools are necessary for identifying the teachers' social-emotional competencies because who has an inadequate level of social-emotional competencies or needs intervention is decided based on the scale scores. Hence, the factor structures of scales, such as scales of emotional awareness and personality factors, have an essential role in the validity of the inferences made from those measurements (Kane, 2013).

The Aim of the Study

The purpose of this study is to provide an empirical illustration of how a joint analysis of the results obtained from different factor analytic model applications on the same dataset can help obtain a better picture of the structure underlying the responses than those that would be obtained from only one model application of choice. In this study, four alternative factor models are considered: 1) the unidimensional model (Dunn & McCary, 2020), 2) the correlated model (Brown, 2015), 3) the higher- (or second-) order model (Thurstone, 1944), and 4) the bifactor model (Holzinger & Swineford, 1937). The datasets were collected from The Emotional Awareness Questionnaire (Rieffe et al., 2008) (Study 1) and the Big Five Personality Test (Goldberg, 1992) (Study 2). The scales, datasets, and methods are described in the following text.

METHOD

Study I

Study I sample comprised 670 preservice teachers (58% of the participants are female and 13% of them are male, gender information of 29% of the sample could not be reached). The data were collected from volunteer preservice teachers in the fall and spring terms of the 2021-2022 academic year.

The Emotional Awareness Questionnaire (EAQ) was used to measure individuals' emotional awareness in Study I. Emotional awareness refers to an attentional process interconnected with some interpretative and evaluative functions (Rieffe et al., 2008). The self-report EAQ aims to identify how people feel and think about their feelings. The EAQ developed by Rieffe et al. (2008), has been adapted into Turkish by Inceman-Kara and Yuksel (2022). The EAQ has 30 items and six sub-factors describing different aspects of emotional functioning: (1) differentiating emotions, (2) verbal sharing of emotions, (3) not hiding emotions, (4) bodily awareness, (5) attending to others, and (6) analyses of emotions. Respondents are asked to rate the degree to which each item is suitable for them on a 5-point Likert scale (from 1 = not true to 5 = always true). Cronbach's α reliability coefficients for each sub-factor were 0.82, 0.71, 0.74, 0.82, 0.82, and 0.81.

Study II

Study II sample comprised 670 preservice teachers (78% of the participants are female, and %22 of them are male). The Big Five Personality Test (BFPT; Goldberg, 1992) was used to measure personality factors in Study II.

The BFPT was developed by Goldberg (1992) and adapted into Turkish by Morsunbul (2014). Personality traits can be assessed under the five sub-factors in the BFPT. These factors are (1) extraversion, (2) agreeableness, (3) emotional stability, (4) conscientiousness, and (5) openness to experience. The BFPT has a total of 30 items, and each sub-factor consists of six items. Items can be responded to on a 7-point Likert scale (from 1 = does not apply to me at all to 7 = applies to me very well). Cronbach α reliability coefficients for each sub-factor were 0.83, 0.78, 0.79, 0.83, and 0.84.

Factor Analytic Models Used in the Applications

Four alternative Confirmatory Factor Analytic (CFA) models were considered: 1) the unidimensional model, 2) the correlated model, 3) the higher- (or second-) order model, and 4) the bifactor model. The CFA models were tested for Study I[1] and Study II.

The Unidimensional Model

Unidimensionality is the main assumption within Classical Test Theory (CTT) and Item Response Theory (IRT) (Embretson & Reise, 2000). The unidimensional model (see Figure 1a) hypothesizes a single latent factor to explain the variance across all observed variables (Dunn & McCary, 2020). Estimated factor loadings indicate the power of the relationship between a single factor and each observed variable. Error terms estimated for each observed variable indicate unexplained variance by a single latent factor. The single factor represents the construct

[1] Since the bifactor model did not converge for Study I, the results of unidimensional, correlated, and higher-order models were reported in findings.

that an instrument intends to measure. (Rodriquez et al., 2016). A key point when using this model is that the test is unidimensional. In practice, many tests or scales may not provide unidimensional assumptions.

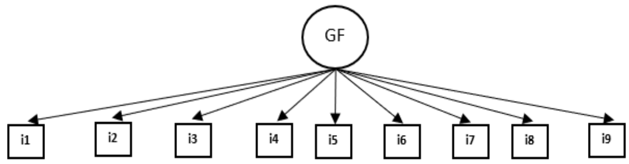
The Correlated Model

The correlated model (see Figure 1b) has generally been used for comparison with the unidimensional model. The correlated model (Brown, 2015) includes two or more latent variables that can correlate. Observed variables are grouped by shared properties and act as indicators of a factor hypothesized to reflect this commonality. This explicitly models the multidimensionality of a test. Estimated factor loadings indicate the power of the relationship between observed variables and their associated factors (Dunn & McCary, 2020). The correlations between latent variables indicate shared variation among latent variables.

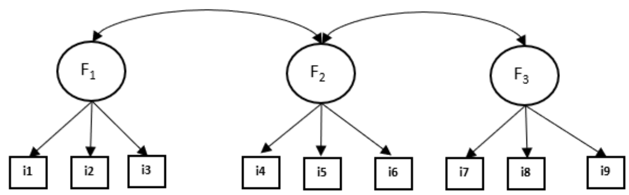
The Higher-Order Model

The higher-order model (Thurstone, 1944) incorporates at least one general (higher-order) factor and a series of sub-factors upon which a specified sub-group of items (see Figure 1c). The higher-order factor models the shared variance between sub-factors (Dunn & McCary, 2020), which means that first-order factors are independent, and each first-order factor mediates the relationship between the higher-order factor and the observed variables. There is no direct effect between higher-order factors and observed variables. The observed variables perform as indicators of the sub-factors, and the commonality modeled by higher-order factors is between the scales already established for each sub-factor (Dunn & McCary, 2020).

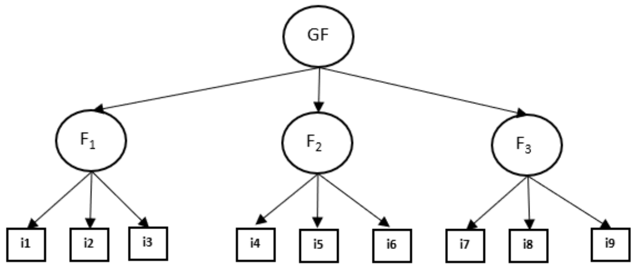
A researcher might seek empirical evidence for reporting an overall score in addition to sub-scores for each sub-domain incorporated in the test or scale. Suppose the loadings between the higher-order and sub-factors are satisfactorily high; in that case, it can be concluded that there is enough commonality between the sub-factors to justify this reporting of both sub-scores and an overall score (Dunn & McCary, 2020).



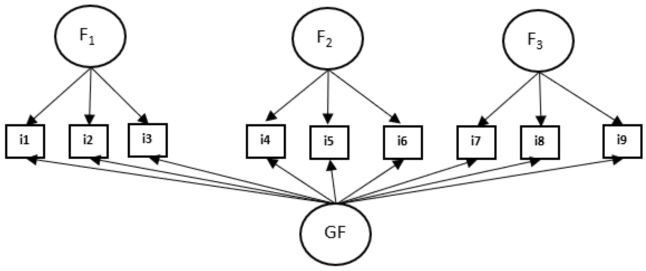
a) Unidimensional model



b) Correlated model



c) Higher-order model



d) Bifactor model

Note. GF=General Factor, F=Factor, i=item

Figure 1. *Four Types of Factor Models*

The Bifactor Model

The bifactor model (Holzinger & Swineford, 1937) specifies that for a given set of item responses, correlations among items can be accounted for by (1) a general factor representing shared variance among all the items and (2) a set of group factors where variance over and above the general factor is shared among subsets of items presumed to be highly similar in content (see Figure 1d) (Rodriguez et al., 2015). The general factor indicates the construct an instrument intends to measure, and group factors indicate more specific subdomain constructs (Rodriguez et al., 2015). The interpretation of the general factor loadings is the same as the single factor in the unidimensional model. The grouping factors estimate the shared variance between sub-groups of items once the common variance between all observed variables captured by the general factor has been partitioned out (Dunn & McCary, 2020). One of the defining features of the bifactor model is that the grouping factors are hypothesized to be orthogonal (uncorrelated) with the general factor and each other (Markon, 2019; Rijmen, 2010). Bifactor models have been used for testing for some situations: (1) to check multidimensionality (Chen et al., 2006), (2) to decide the appropriateness of a total score and what, if anything, one might gain by scoring subscales (Reise, 2012). The bifactor model can help the researcher to decide whether the test or scale is unidimensional enough to be reported on a single factor or whether it makes sense to report domain sub-scores.

In addition to analyze factor analytic models, reliability indices were derived for the models. These indices include Cronbach's alpha (α) (Cronbach, 1951) and McDonald's omega (ω) coefficients (Mcdonald, 1999). The omega coefficient is computed for composite score reliability. The variance of the general and group factors are combined to obtain the reliability estimate for the reliability of total and subscale scores. The differences between coefficients alpha and omega are that (a) omega always is based on the factor loadings of a specific model, whereas alpha, typically, is computed based on observed variances and covariances, and (b) alpha assumes equal loadings, whereas omega is more appropriate when factor loadings vary (Reise, 2012).

In summary, the unidimensional and bifactor models directly model shared variance between observed variables, while the correlated factors and higher-order model mediate this relationship by the inclusion of grouping factors at the first-order level (Dunn & McCary, 2020). These varying structures derive the researcher's different perceptions of the measurement properties of a scale. The four CFA models, unidimensional, correlated, higher- (or second) order, and bifactor, were fitted to the two datasets using the latent variable modeling software M*plus* (version 8.3) (Muthén & Muthén, 1998-2017) with Maximum Likelihood with Robust Standard Errors (MLR) estimation method.

The model χ^2 values and associated *p*-values were reported for model evaluation. Other model fit indices were also reported because χ^2 values are sensitive to sample size. RMSEA and SRMR (<.08), and CFI/TLI (>.90) values are taken as references in model comparisons (Hu & Bentler, 1999). Since the models are not all nested, information criteria values - AIC, BIC, and SABIC - were reported to compare the fit of non-nested models. Lower information criteria values indicate better model-data fit.

Research Ethics

All ethical procedures were completed in this study. Ethical permission for the research was approved by Gazi University Ethics Committee. The ethics committee document number is 7082166-604.01.02.

FINDINGS

Study I Findings

The model-fit indices of the EAQ data for the three CFA models are shown in Table 1. For all models, χ^2 *p*-values were statistically significant ($p < .05$). In terms of statistical measures of comparative fit, the best fit was achieved by the higher-order factor model, followed by the correlated model, and the unidimensional model yielded the worst fit. The average factor loadings on the unidimensional model were 0.31, indicating a mean explanation of 10% of the variance of the observed variables. Some of the observed variable factor loadings were less than 0.30. As a rule of thumb, variables with loadings of 0.32 and above are interpreted (Brown, 2015; Tabachnick & Fidell, 2013). This means that the indicator is meaningfully related to a primary or secondary factor. Therefore, unidimensional model factor loadings seem not to ensure this rule.

The average factor loadings for the correlated model were 0.30 and above. In the higher-order model, all first-order factors loaded medium level on the general factor. The average of the item loadings on the first-order factors was between 0.59 and 0.75, indicating that there was evidence of multidimensionality for the construct of the EAQ.

Table 1. Model-Data Fit Indices of Four Models for the EAQ

	Model-data fit indices		
Model-fit	Unidimensional	Correlated	Higher-order*
χ^2 (df)	5047.19 (405)	1333.55 (390)	1352.9 (397)
AIC	56282.51	52598.88	52604.23
BIC	56688.17	53072.14	53045.94
SABIC	56402.41	52738.76	52734.78
RMSEA	0.13	0.06	0.06
SRMR	0.13	0.06	0.08
CFI	0.39	0.88	0.88
TLI	0.35	0.86	0.86
Mean (SD) of factor loadings			
General	0.31 (0.31)	-	0.49 (0.34)
F_1	-	0.40 (0.16)	0.62 (0.12)
F_2	-	0.30 (0.18)	0.75 (0.05)
F_3	-	0.47 (0.08)	0.59 (0.15)
F_4	-	0.45 (0.15)	0.68 (0.13)
F_5	-	0.38 (0.21)	0.70 (0.07)
F_6	-	0.37 (0.18)	0.70 (0.11)
Factor correlations			
$F_1 – F_2$	-	0.55	-
$F_2 – F_3$	-	0.70	-
$F_5 – F_6$	-	-0.64	-

*Retained model.

The lower AIC presented the correlated model compared to the unidimensional model to provide a more accurate description of the EAQ data. Factor correlations above 0.80 may imply poor discriminant validity (Brown, 2015). Factor correlations in the model were observed to be both positive and negative; less than 0.80 means good discriminant validity. The estimated factor loadings on the correlated model were an average of 0.30 and above. CFI (0.88) and TLI (0.86) indices showed that the correlated model had a better solution than the unidimensional model. Brown (2015) suggests that estimating factor correlations provides essential information, such as the existence of redundant factors or a potential higher-order structure. Estimation of the higher-order factor model could provide additional insight.

The comparative fit of the higher-order model was favorable to the correlated model, with a lower value of BIC and SABIC. Both models had CFI and TLI values close to the suggested threshold of 0.90, indicating acceptable levels of fit. Similarly, the RMSEA and SRMR were also acceptable levels of fit for both models. However, given the equally global fit statistics, the higher-order model would be accepted to understand the factor structure. The factor loadings of the higher-order factor on the six sub-factors provided a good summary of them, as seen in Figure 2a. These results indicate that the EAQ is multidimensional.

The overall emotional awareness score was represented by the general factor, while item-group-specific factor scores were represented by the first-order factors. The general factor had reliability coefficients of Cronbach's $\alpha = 0.55$ and McDonald's $\omega = 0.63$. For the six sub-factors, the reliability coefficients were McDonald's ω = 0.83, 0.80, 0.74, 0.82, 0.64, and 0.83, respectively.

Study II Findings

Table 2 displays the model-fit indices for the four models of BFPT data. It can be seen that all models' χ^2 p-values were statistically significant ($p < .05$). The bifactor model had the best fit according to statistical measures of model-data fit, followed by the higher-order model, the correlated model, and the unidimensional model, which had the poorest fit.

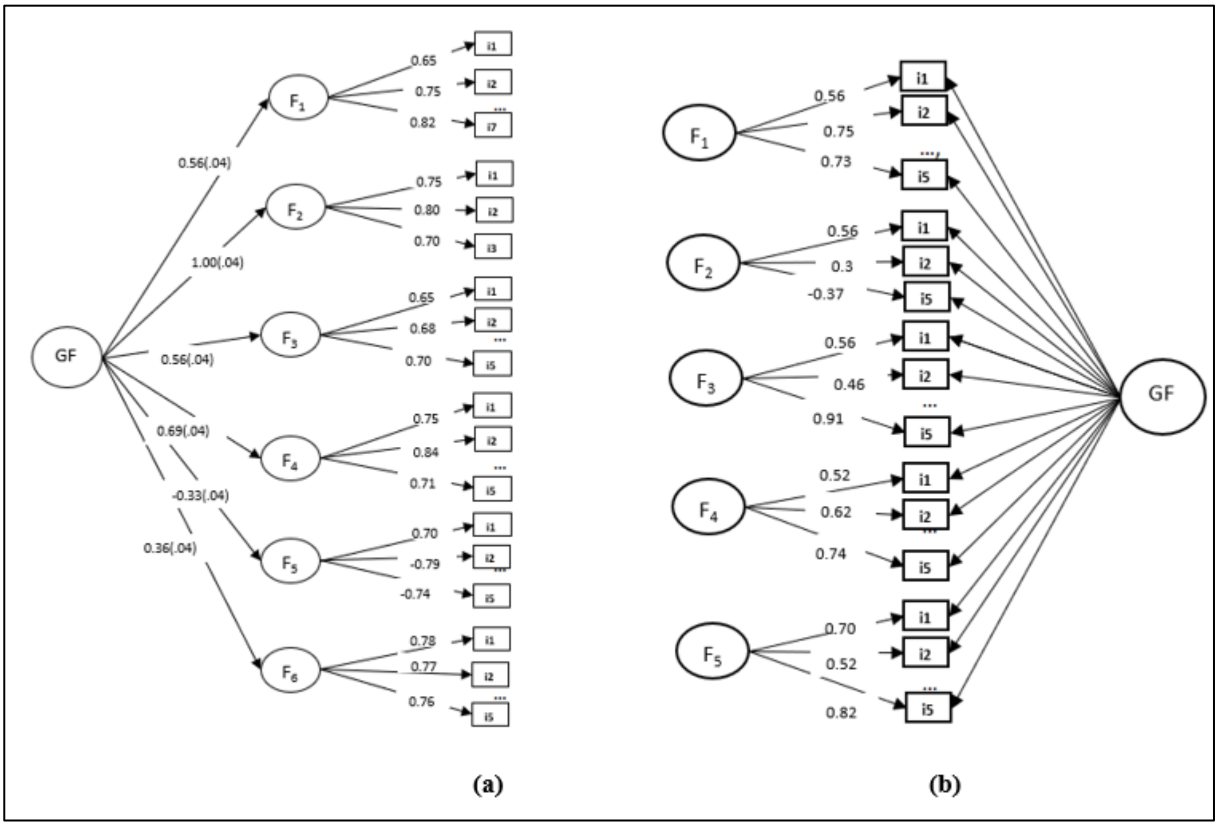
Table 2. Model-Data Fit Indices of Four Models for the BFPT

	Model-data fit indices			
Model-fit	Unidimensional	Correlated	Higher-order	Bifactor*
χ^2 *(df)*	7208.46 (405)	2195.46 (395)	2335.94 (400)	1973.38 (381)
AIC	70648.89	65655.89	65786.37	65461.81
BIC	71054.55	66106.62	66214.56	65975.64
SABIC	70768.79	65789.11	65912.93	65613.68
RMSEA	0.16	0.08	0.08	0.07
SRMR	0.15	0.08	0.09	0.12
CFI	0.25	0.80	0.79	0.83
TLI	0.19	0.78	0.77	0.80
Mean (SD) of factor loadings				
General	0.39 (0.10)	-	0.36 (0.21)	0.29 (0.26)
F_1	-	0.64 (0.10)	0.64 (0.09)	0.65 (0.10)
F_2	-	0.63 (0.26)	0.63 (0.26)	0.36 (0.11)
F_3	-	0.69 (0.19)	0.69 (0.19)	0.65 (0.20)
F_4	-	0.62 (0.22)	0.62 (0.22)	0.60 (0.14)
F_5	-	0.69 (0.11)	0.69 (0.11)	0.66 (0.12)
Factor correlations				
F_1 – F_5	-	0.37	-	-
F_1 – F_3	-	0.31	-	-
F_2 – F_4	-	0.46	-	-

*Retained model.

The average factor loadings on the unidimensional model were 0.39, indicating a mean explanation of 16% of the variance of the observed variables. The observed variables were related to a primary factor in the unidimensional model. For the correlated model, the estimated factor loadings were an average of 0.62 and above. The average of the item loadings for the first-order factors in the higher-order model was the same as those in the correlated model. In the bifactor model, the average loading on the general factor (0.29) was close to that of the unidimensional model (0.39); however, the mean loadings for the grouping factors were between 0.36 and 0.66, indicating that there might be evidence of multidimensionality for the construct of interest.

The lower AIC and BIC presented the correlated model compared to the unidimensional model to provide an accurate description of the BFPT data. Comparative fit indices of the correlated model -CFI (0.80) and TLI (0.78)- indicated a better solution to the data. It seemed to be a better model improvement. Factor correlations in the model were both positive and negative; less than 0.80 implied good discriminant validity. The comparative fit of the higher-order model was not favorable to the correlated model, with a higher score of AIC, BIC, and SABIC. The higher-order model was far from the suggested threshold of 0.90 for CFI and TLI. The correlated and higher-order models had acceptable levels of fit on the RMSEA and SRMR. Further investigation of multidimensionality, the data were modeled using the bifactor model.



Note. a = Higher-order model for the EAQ; b = Bifactor model for the BFPT.

Figure 2. *Final Model Diagrams for the EAQ and BFPT Data*

The bifactor model demonstrated a better solution than the higher-order factor model, with lower AIC, BIC, and SABIC values. The bifactor model was the closest to the threshold of 0.90 for CFI and TLI. Additionally, the bifactor model showed better RMSEA value compared to the higher-order model. Therefore, the bifactor model was chosen to understand the BFPT factor structure. In the bifactor model, the grouping factor estimates indicated the degree of shared variance between group items after accounting for the general factor (Dunn & McCary, 2020). For the first grouping factor (F1), item loadings were greater than 0.50, with a mean of 0.65 which means that the grouping factor explains more than 40% of the observed variance across items associated with the first factor. It indicated a systematic deviation from the variance explained by the general factor. Overall, the mean loadings on the other grouping factors were 0.36, 0.65, 0.60, and 0.66, which indicates that, separately, around 40% of the observed variance in each group of items could be explained by grouping factors. These findings provided some evidence of the multidimensionality of the BFPT. Figure 2b gives factor loadings diagrams for the bifactor model.

The general factor indicated overall personality trait scores; grouping factors indicated item-specific factor scores. The reliability coefficients of the general factor were Cronbach's $\alpha = 0.83$ and McDonald's $\omega = 0.85$. The reliability coefficients of the first sub-factor were McDonald's $\omega = 0.81$; 0.84 for the second sub-factor; 0.85 for the third sub-factor; 0.80 for the fourth sub-factor, and 0.85 for the fifth sub-factor

DISCUSSION AND CONCLUSION

The results show that the simultaneous use of alternative factor analytic models can potentially provide invaluable information that might be overlooked if only one model were to be used when investigating the dimensional structures of the constructs measured. The two illustrative examples provided in this paper show that such an investigative approach might be especially suited for studies involving measurements of social-emotional competencies. The illustrative examples demonstrate that employing different methodologies can significantly improve both the conclusions drawn about dimensionality structures and the interpretations of the measurements. In this study, the results of the unidimensional model were not suited for both scales due to the multidimensional structures. Therefore, we preferred the multidimensional CFA models to test the structures of the scales. The first study provided evidence of the multidimensionality of the EAQ, as demonstrated by the fit statistics of the higher-order model. The results of the higher-order factor model supported the idea that the EAQ had a multidimensional construct, indicating that first-order and higher-order scores could be reported separately. The second study examined the dimensionality of the BFPT factors. Again, we got some evidence of the multidimensionality of the BFPT. The fit of the bifactor model indicated the need to report sub-scores for personality traits (i.e., extraversion, agreeableness, and so on) and general factor score. The results show that different CFA models were better suited for the multidimensional structures measured by the two studies.

The identification of multidimensionality is frequently used as both necessary and sufficient justification for reporting and scoring subscale scores (Reise et al., 2013). Selecting the appropriate measurement model is important for both accurately reporting scores and evaluating the reliability of those scores (Brunner et al., 2012). Total scores are typically more reliable as they are based on a larger number of items, whereas subscale scores, which are used when the construct of interest is multidimensional, may be based on a smaller number of items and may, therefore, be somewhat less reliable. Nonetheless, sub-scores are often sought after by the stakeholders due to the common perception that (a) sub-scores would provide credible information about the examinee's strengths and weaknesses, and (b) the examinee would work harder on the categories on which he/she performed poorly and hence, might improve in those areas (Sinharay et al., 2011). For these perceptions, at least reasonably, to be accurate, it is highly recommended that evidence of adequate sub-scale score reliability and validity arguments must be established and presented. In the reliability context, the findings showing a better fit for the higher-order model for the EAQ scale support that the EAQ sub-scale scores could be preferred considering their higher reliability compared to those of the total scores. For the BFPT data, however, the bifactor model findings support that the sub-scale scores and general factor scores are comparable in their reliabilities, and, hence, the total score and the subscale scores could be reported. The group factors in the bifactor model (in this case, extraversion, agreeableness, etc.) were represented by a common source of variance, controlling for the common variance explained by the general factor, i.e., the personality trait. A similar finding was also reported by Nguyen and Biderman’s (2013) studies associated with BFPT dimensionality.

Although many educational and psychological measures are designed primarily to scale individuals on a single construct, many psychological traits (e.g., social-emotional competencies) often are theoretically defined to have sub-traits (corresponding measures to be captured by several items) (Rodriguez et al., 2015). Consequently, many common psychological measurement tools (tests or scales) consisting of multiple items can be suspected to be multidimensional, at least to some extent. Hence, social science researchers are recommended to initiate their own investigations into the structure underlying response data they collected using tests or scales that might be multidimensional to some extent (Gustafsson & Åberg-Bengtsson, 2010). At this point, it is crucial that the proper statistical modeling tools are used for the intended score interpretations (Brunner et al., 2012).

Some CFA models help the evidence-gathering process for the multidimensionality of the construct. The bifactor model allows the researcher to investigate the assumption of a combination of general and domain-specific abilities (Dunn & McCary, 2020). It is stated that the bifactor model can be understood as complementary to the higher-order model. In the higher-order model, superordinate factors influence subordinate factors, but in the bifactor model, superordinate factors influence observed variables (Markon, 2019). Gustafsson and Åberg-Bengtsson (2010) highlight that both models share two types of factors, and the difference between models lies in whether a simple or complex structure is retained. The bifactor model is more complicated as a latent structure than the higher-order model.

Limitations

This paper mainly conducted CFA model analyses to investigate social-emotional competencies’ factor structures. The CFA techniques presented here are relevant to any measure where researchers debate its dimensional structure and ask whether total or subscale scores should be reported or used in their research. Albeit limited to four models and two data sets, the findings might be useful to researchers who are interested in alternative ways of studying scale dimensionality. Through the illustrative examples, each of which came to substantively different conclusions about the dimensionality of the test or scale, it was hoped that an example for the usage of the alternative factor analytic modeling in psychological assessment research had been provided, and recommendations have been given on how to approach inference from the model..

Statements of Publication Ethics

This research study complies with research and publishing ethics. The studies involving human participants are reviewed and approved by the Gazi University Ethics Committee (Ethics Committee Number: 7082166-604.01.02)..

Acknowledgments

This study was presented at Lisansüstü Öğretmen Çalışmaları Kongresi (13-15 May 2022), Balıkesir, Türkiye. This study was partially supported by the Gazi University Scientific Research Council [grant number 04/2019-01].

Researchers' Contribution Rate

Nilufer Kahraman: Data Collection, Method, Supervision and Conclusion. *Esra Sozer-Boz***:** Literatur Review, Data Collection, Data Analysis, Results, and Conclusion. *Derya Akbas:* Literatur Review, Data Collection and Conclusion. *Ergün Cihat Çorbacı:* Data Collection. *Şerife Işık:* Data Collection. *Nazife Üzbe Atalay:* Data Collection. *Fatma Nur Aydın:* Data Collection. *Mehtap Çakan:* Literatur Review. *Şeref Sağıroğlu:*Data Collection

Conflict of Interest

The authors declare no conflict of interest.

REFERENCES

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford.

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality, 80*, 796-846. https://doi.org/10.1111/j.1467-6494.2011.00749.x

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105. https://doi.org/10.1037/h0046016

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of Bifactor and Second-order models of quality of life. *Multivariate Behavioral Research, 41*(2), 189-225. https://doi.org/10.1207/s15327906mbr4102_5

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334. https://doi.org/10.1007/BF02310555

Dunn, K. J., & McCary, G. (2020). The place of the bifactor model in confirmatory factor analysis investigations into construct dimensionality in Language Testing. *Front. Psychol.*, 1357. https://doi.org/10.3389/fpsyg.2020.01357

Eid, M., & Diener, E. (2006). Introduction: The need for multimethod measurement in psychology. *In M. E. Diener, Handbook of multimethod measurement in psychology* (pp. 3-9). American Psychological Association.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*(1), 26-42. https://doi.org/10.1037/1040-3590.4.1.26

Gustafsson, J. E., & Åberg-Bengtsson, L. (2010). Unidimensionality and the interpretability of psychological instruments. *In S. Embretson, Measuring psychological constructs: Advances in model-based approaches* (pp. 97-121). American Psychological Association. https://doi.org/10.1037/12074-005

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika, 2*, 41-54. https://doi.org/10.1007/BF02287965

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

Inceman-Kara, F., & Yuksel, G. (2022). Focusing on Adolescents' Emotional Awareness: Instrument Validation and Evaluation of a Training Program. *Participatory Educational Research, 9*(4), 289-306. http://dx.doi.org/10.17275/per.22.91.9.4

Jennings, P. A., & Greenberg, M. T. (2009). The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research, 79*, 491-525. http://dx.doi.org/10.3102/0034654308325693

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73. https://doi.org/10.1111/jedm.12000

Markon, K. E. (2019). Bifactor and hierarchical models: specification, inference, and interpretation. *Annual Review of Clinical Psychology, 15*, 51-69. https://doi.org/10.1146/annurevclinpsy-050718-095522

McCarthy, D. (2021). Adding social emotional awareness to teacher education field experiences. *The Teacher Educator, 56*(3), 287-304. https://doi.org/10.1080/08878730.2021.1890291

McDonald, R. P. (1999). *Test theory: A unified approach*. Erlbaum.

Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749. https://doi.org/10.1037/0003-066X.50.9.741

Morsunbul, (2014). The validity and reliability study of the Turkish version of quick big five personality test. *Dusunen Adam, 27*(4), 316-322. https://doi.org/10.5350/DAJPN2014270405

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

Nyugen, N. T., & Biderman, M. (2013). Predicting Counterproductive Work Behavior from a Bi-factor Model of Big Five Personality. *Academy of Management Annual Meeting Proceedings*. https://doi.org/10.5465/ambpp.2013.13719abstract

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*, 667-696. http://dx.doi.org/101080/00273171.2012.715555

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*, 129-140. http://dx.doi.org/10.1080/00223891.2012.725437

Rieffe, C., Oosterveld, P., Miers, A. C., Terwogt, M. M., & & Ly, V. (2008). Emotion awareness and internalising symptoms in children and adolescents: The Emotion Awareness Questionnaire revised. *Personality and Individual Differences, 45*(8), 756-761. https://doi.org/10.1016/j.paid.2008.08.001

Rijmen, F. (2010). Formal Relations and an Empirical Comparison among the Bi-Factor, the Testlet, and a Second-Order Multi-dimensional IRT Model. *Journal of Educational Measurement, 47*(3), 361-372. https://doi.org/10.1111/j.1745-3984.2010.00118.x

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2015). Evaluating bifactor models: calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137-150. http://dx.doi.org/10.1037/met0000045

Rodriquez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactorsStatistical indices in the evaluation of psychological measures. *Journal of Personality Assessment, 98*(3), 223-237. http://dx.doi.org/10.1080/00223891.2015.1089249

Sinharay, S., Puhan, G., & Haberman, S. J. (2011). An NCME instructional module on subscores. *Educational Measurement: Issues and Practice, 30*, 29-40. https://doi.org/10.1111/j.1745-3992.2011.00208.x

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Pearson.

Thurstone, L. (1944). Second-order factors. *Psychometrika, 9*(2), 71-100. http://dx.doi.org/10.1007/bf02288715

Ulloa, M., Evans, I., & Jones, L. (2016). The effects of emotional awareness training on teachers' ability to manage the emotions of preschool children: An experimental study. *Escritos de Psicología, 9*(1), 1-14. http://dx.doi.org/10.5231/psy.writ.2015.1711