Research Methods in Applied Linguistics 1 (2022) 100005

Contents lists available at ScienceDirect

Research Methods in Applied Linguistics

journal homepage: www.elsevier.com/locate/rmal

Exploratory structural equation modeling (ESEM) and bifactor ESEM for construct validation purposes: Guidelines and applied example

Abdullah Alamer[a,b]

[a] Department of English, Al-Ahsa Campus, Imam Mohammad Ibn Saud Islamic University (IMSIU), 9134 Naseem, Hofuf, Western 36363, Saudi Arabia
[b] Department of English, King Faisal University

ARTICLE INFO

Keywords:
Exploratory structural equation modeling (ESEM)
Confirmatory factor analysis (CFA)
Basic psychological needs
construct validity
Convergent and discriminant validity
Factor analysis

ABSTRACT

The present study aims to introduce the advancement of the exploratory structural equation modeling (ESEM) and bifactor ESEM for construct validation purposes in second language research. It does so by applying the methods to the Basic Psychological Needs in Second Language (BPN-L2) Scale (Alamer, 2021a) to assess its factorial validity from a novel perspective. Bifactor ESEM is useful in situations where the theory postulates two layers of constructs (global and specific constructs). A comparison was made between four CFA and ESEM models, and the bifactor ESEM was the model of choice because it provided sufficient details about the multidimensionality of the BPN-L2 scale which accounted for the specific constructs (i.e., autonomy, competence, and relatedness) and the general construct (i.e., the basic psychological needs) and allowed for the assessment of convergent, discriminant, and predictive (criterion) validity from a new perspective. Predictive (criterion-related) validity was assessed using the structural bifactor ESEM by showing that the specific factors explained autonomous motivation differently compared to the global BPN factor, thus confirming the presence of *imbalance* at the specific factors level. The study provided guidelines in applying bifactor ESEM and presented an applied example with the syntax required to run the models in Mplus. The findings showed that the BPN-L2 scale is reliable and valid through the bifactor ESEM solution and that CFA solutions failed to support the measurement model.

Introduction

Quantitative researchers in second language (L2) research use different types of measures to assess language learners' profiles from psychological, motivational, behavioral perspectives (Woodrow, 2014), and researchers are required to check the extent to which their measures are valid. Validity is multifaced and can be evaluated through different types of assessments, and one important type of validity is *construct* validity (Morin et al., 2016). Construct validity is an essential component in L2 quantitative research to determine statistically the appropriateness of the measure (also called instrument, scale and questionnaire) and whether it measures what is supposed to measure. Researchers usually rely on different factor analysis methods to obtain information about how the items (i.e., observed variables) correspond to their factors (i.e., latent variables) in the measure. Hence, factor analysis is used to examine the internal structure of the constructs, which is one aspect of the construct validity. The dominant factor analysis methods used in the field are possibly exploratory factor analysis (EFA) and principal components analysis (PCA) followed by confirmatory factor analysis (CFA) (Plonsky, 2015; Woodrow, 2014). However, investigating the underlying factor structure of the scales is a critical task

E-mail address: alamer.aaa@gmail.com

https://doi.org/10.1016/j.rmal.2022.100005
Received 10 October 2021; Received in revised form 28 January 2022; Accepted 30 January 2022

and needs to be carried out using the most appropriate techniques. The field can benefit from recent developments in psychometric multidimensionality research to gain precise and more accurate results of the measures in question.

Specifically, recent advancements in psychometric research brought a new method called exploratory structural equation modeling (ESEM) that integrates the best of CFA and EFA in one analytical framework (Alamer & Marsh, 2022; Asparouhov & Muthén, 2009; Marsh et al., 2009; Morin et al., 2013). ESEM is similar to EFA as it allows items to cross-load on all factors involved in the analysis, but it differs from the EFA as it takes on the features of structural equation modeling (SEM) such as evaluating model fit indices, assessing the measurement errors, testing model invariance and allowing different model specifications to be included. Within the ESEM and CFA models, a bifactor configuration can be estimated (also called nested-factor model or general-specific model). Bifactor models are viable solutions when the theory specifies the existence of global (also called general) factors along with their specific factors (Alamer, 2021b; Chen et al., 2006; Kline, 2016; Sánchez-Oliva et al., 2017). In this study, I present to the readers the advancement of the bifactor ESEM by applying it to the Basic Psychological Needs in Second Language (BPN-L2) Scale (Alamer, 2021a) and testing its dimensionality. Because the BPN-L2 scale assumes specific and general factors in one measurement model (explained in a later section), neither standard CFA nor standard ESEM can capture the general construct. Hence, the bifactor ESEM (and bifactor CFA if supported) can be a suitable alternative solution.

The need for exploratory structural equation modelling (ESEM)

In essence, EFA involves exploring data in search of possible latent variables (or factors) based on observed variables (or items) involved in the data. Therefore, EFA and PCA are usually conducted when researchers do not have clear information about the factor structure of their data (Alamer & Marsh, 2022; Plonsky, 2014, 2015; Roever & Phakiti, 2017). However, experts maintained that the validity of the constructs cannot be fully evaluated by solely using EFA or PCA (Alamer, 2021b, Alamer, 2021c; Alamer & Marsh, 2022; Hair et al., 2019; Marsh et al., 2009; Woodrow, 2014). This is because of the limitations EFA and PCA have compared to CFA such as the lack of the goodness-of-fit indices and the inability to compare the fit of competing models, and across different samples (test of invariance). Methodologically PCA is perhaps unsuitable when it is used to confirm measuremnt models because its main objective is to reduce the number of items (Hair et al., 2019). A key characteristic of the CFA (which is also a drawback) is that items load *only* on the factors they are presumed to belong to. In this way, cross-loadings across other relevant factors are specified to *zero*. Accordingly, researchers prefer CFA when they have a hypothesized structure of the factors that exist in the data (i.e., theory-driven assessment). But a CFA model with such independent factor loadings requires solid theoretical and empirical justification because conceptually related items usually have some levels of construct-relevant associations with the nontarget construct(s) in the model. As such, there are drawbacks to overlooking cross-loadings in assessing measurement models when the constructs have certain levels of conceptual similarities. Ignoring cross-loadings often leads to biased results regarding parameter estimates such as inflated factor correlations and reduced goodness-of-fit indices (Alamer & Marsh, 2022; Marsh et al., 2009, 2020). This is because items do not perfectly represent only their corresponding factor; rather, they can overlap and associate with other constructs involved in the assessment that are conceptually related (Gillet et al., 2020; Guay et al., 2015; Morin et al., 2020). Accordingly, relying on CFA might result in construct misspecification due to the inflation of the correlations between the factors, which substantiate concerns regarding the discriminant validity of the constructs. Because CFA is restricted in nature (Cromhout et al., 2021; Marsh et al., 2009, 2014) researchers are often confronted with ill-fitted models leading them to apply different post hoc strategies such as correlating multiple measurement errors (uniqueness) or removing several items from the factors to compensate for the inappropriately specified CFA model. But such strategies rarely make theoretical sense (i.e., they only provide empirical support). For example, correlating multiple measurement errors imply that there is an underlying (i.e., confounding) factor that the researcher failed to account for in the assessment. Similarly, removing items from the measure should not be always preferred to improve the fit of the model especially when the factor is defined by a limited number of items (such as three). With a limited number of items, each item carries important information about the meaning of the construct that, when removed, would result in a distorted representation of the concept (Hair et al., 2019). Examples from L2 research that highlighted these issues can be found in Alamer and Marsh (2022) and Alamer (2021b).

This is where ESEM comes into play. ESEM has been developed as an improved method that overcomes the limitations of EFA and CFA and integrates the best of the two worlds (Howard et al., 2018; Morin et al., 2013). In this way, ESEM simultaneously allows cross-loadings among items and factors (which is an EFA feature) but also computes goodness-of-fit indices, permits error terms to be estimated, and allows for invariance test (which are CFA features).

In almost all multidimensional studies, ESEM provides a better fit to the data because of the flexibility and less restrictive system it has compared with CFA (see Morin et al., 2020 for a review). Furthermore, researchers note that forcing conceptually related factor loadings to be zero results in biased estimation of the factor correlation (Marsh et al., 2014; 2020; Morin et al., 2013; Sánchez-Oliva et al., 2017). When cross-loadings are estimated, even if they are small and non-significant, factor correlations tend to be unbiased and properly reflect the data (Marsh et al., 2014). For these reasons, the fit indices of ESEM tend to be better than CFA.

The study conducted by Alamer and Lee (2019) is an example of the issues that CFA yields in research considering the measurement model of BPN and many other constructs. The study tested the factorial validity of several psychological constructs through multiple CFAs, and among these was the BPN. The initial model included three factors with each having four items. However, the researchers found that the model did not converge. The standardized residual covariances suggested that one item in 'relatedness' is rather problematic in the measurement. Adding error covariance did not offer any improvement. The researchers had to remove this item from the measurement and subsequent analyses. However, as explained earlier, the removed item may have important information for the factor and can function properly in other model specifications such as in ESEM and bifactor ESEM. Hence, ESEM and bifactor ESEM can help maintain fuller information about the scale by retaining the items in the measurement (Morin et al., 2019).

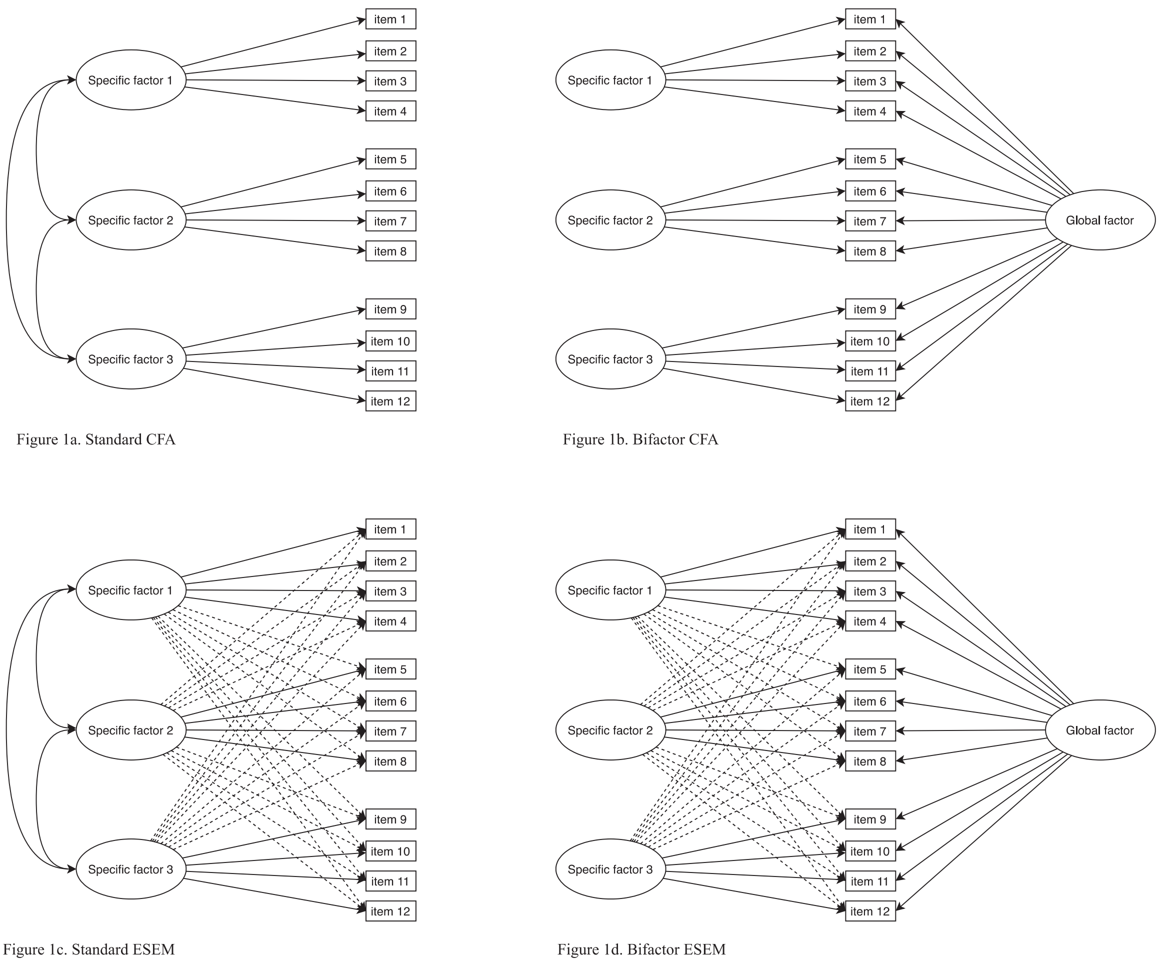

Fig. 1. Visual representation of the four models tested in the present study.

Bifactor models vs. higher-order models

Within the CFA and ESEM frameworks, a bifactor model can be evaluated (Alamer, 2021b, 2021c; Morin et al., 2016, 2020; Reise et al., 2010). Both bifactor and higher-order models postulate the co-existence of the general factors along with specific factors. A unique feature of bifactor models over the higher-order models is that the general factors have direct effects on the indicators (items). Therefore, all items are loaded on one specific factor and one general factor. Consequently, the variance of the indicators is partitioned into three sources: (i) the specific factor, (ii) the general factor, and (iii) the measurement error. In contrast, in the higher-order models, the second-order factor has no direct effect on the indicators but only indirectly through the first-order factors. Thus, several researchers argue against the use of higher-order models in construct-relevant psychometric multidimensionality (Chen et al., 2006; Howard et al., 2018; Marsh et al., 2020; Morin et al., 2016; Sánchez-Oliva et al., 2017). For instance, Howard et al. (2018) stress that bifactor models should be selected over the higher-order models unless strong conceptual justifications are present. Thus, researchers are recommended to retain the bifactor models unless there is a strong theoretical justification for a higher-order representation (Alamer, 2021c; Gillet et al., 2020; Howard et al. 2018). Accordingly, the present study takes this perspective in assessing hierarchically ordered measurement models and does not consider higher-order models in this study.

In bifactor CFA, the items are loaded on their specific factors and the general factor, while in bifactor ESEM all items are loaded on all specific factors and the general factor. In both bifactor CFA and bifactor ESEM the factors are set to be orthogonal (i.e., correlations are set to be zero) (Morin et al., 2020; Tóth-Király et al., 2017) Fig. 1. illustrates the difference between the four models visually. I use the Basic Psychological Needs in Second Language (BPN-L2) Scale (Alamer, 2021a) as an example to show the usefulness and flexibility of the bifactor ESEM compared to bifactor CFA.

Bifactor ESEM to examine the imbalance between the global and specific factors

In many cases, researchers are confronted with situations that require the assessment of two layers (or levels) of the constructs simultaneously. That is, to evaluate the extent to which the specific factors (e.g., the three needs of BPN) as well as the global factor

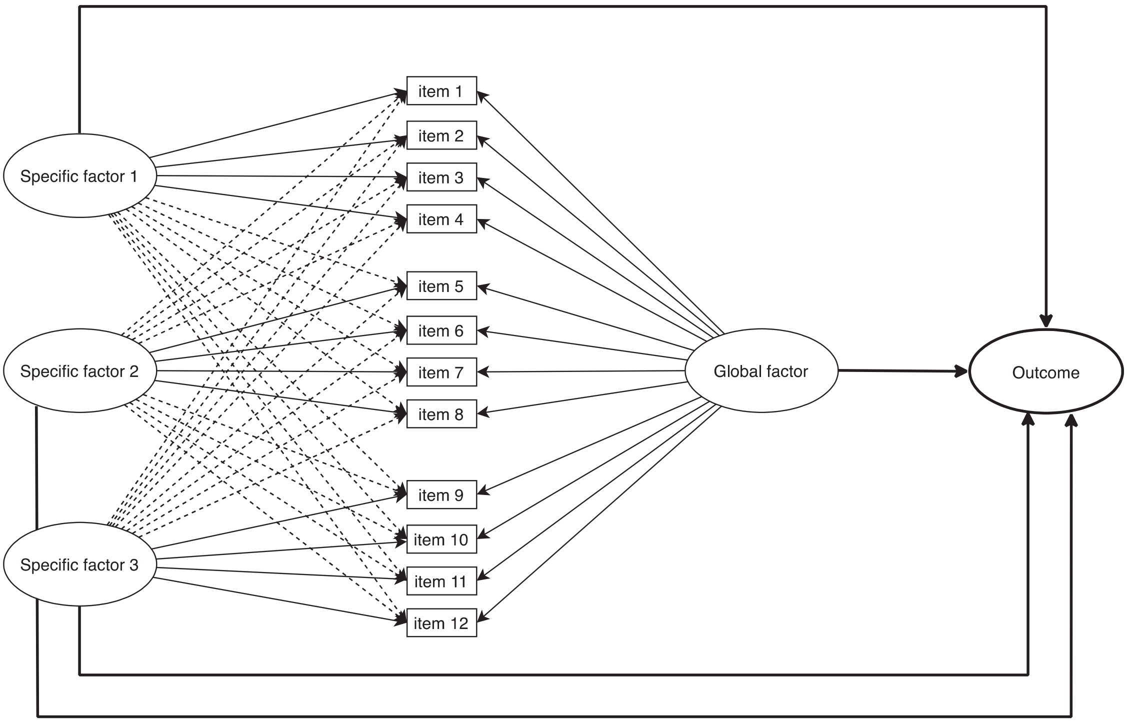

Fig. 2. Visual representation of the structural bifactor ESEM model.

(the general BPN) relate to the language learning outcome. In typical CFA and ESEM models, researchers are only capable of assessing the explanatory power of the first layer of the factors while assuming that the higher (i.e., global) factor layer is achieved implicitly by obtaining the coefficients of the specific factors collectively (some would calculate the average value of the coefficients to argue for the global effect, which is not an unbiased practice). Nonetheless, recent psychometric research illustrated that it is possible to simultaneously consider the two layers using the bifactor models (Alamer, 2021b; Gillet et al., 2020; Sánchez-Oliva et al., 2017). This is an important consideration because research shows that ignoring the two layers in the prediction, when they are assumed to co-exist, carries the risk of missing information about the unique contribution of each specific factor over and above that of the global one. Ignoring the assessment of global factors in the structural model may inflate the magnitude of the coefficients of the specific factors (Tóth-Király et al., 2017) and leave the reader with missing details about the general factor. In the present study, I show the readers how to achieve this and provide evidence about the importance of considering the two levels simultaneously in the assessment of criterion-related validity Fig. 2. shows an illustration of the structural bifactor ESEM model.

When bifactor ESEM should be used and how?

Although its name may suggest an exploration procedure, ESEM's main objective is to test theoretical underpinnings that have been already established. Accordingly, bifactor ESEM should be used when the theory supports the conceptualization of two layers of constructs. A good example is perhaps the specification BPN theory postulates about the global and specific factors (explained in the following section). When the theory supports such conceptualization, bifactor CFA and bifactor ESEM can be included and tested (Tóth-Király et al., 2017). Researchers can use the sequential strategy to compare four alternative models (Morin et al., 2020). That is, the researcher should first evaluate the standard CFA and standard ESEM. If both solutions provided an adequate level of fit to the data, then the researcher should compare the standard solutions to their bifactor counterparts (i.e., comparing four models). If CFA provided similar results to the ESEM in terms of model fit and parameters (e.g., factor loadings and correlations) the researcher should pursue the more parsimonious bifactor CFA model. However, if only standard ESEM provided an acceptable fit to the data compared to the standard CFA, then the researcher should only pursue bifactor ESEM. A decision tree provided by Alamer & Marsh (2022) is used here to guide the researcher select between CFA and ESEM (see Fig. 3 for the decision tree). The selection of the final model should not only be made by empirical results but should be theoretically justified. According to Fig. 3, we can see situations where the researcher may retain the CFA solution. So, the recommendation is not to ignore the inclusion of CFA in the assessment.

It is recommended in ESEM studies to consider the 'target rotation' method to accommodate the confirmatory approach to ESEM. In target rotation, all cross-loadings are 'targeted' to be close to zero, whereas all of the main loadings are freely estimated (Marsh et al., 2014; Morin et al., 2016, 2020). Thus, contrary to some claims, ESEM is a theory-driven approach and is mainly used for confirmatory purposes (Alamer & Marsh, 2022). The guidelines to support the ESEM can be found from Morin et al. (2020), according to which the researcher should prefer the ESEM when (1) the fit indices are improved (e.g., ΔCFI and ΔTLI increase at least by .01, ΔRMSEA

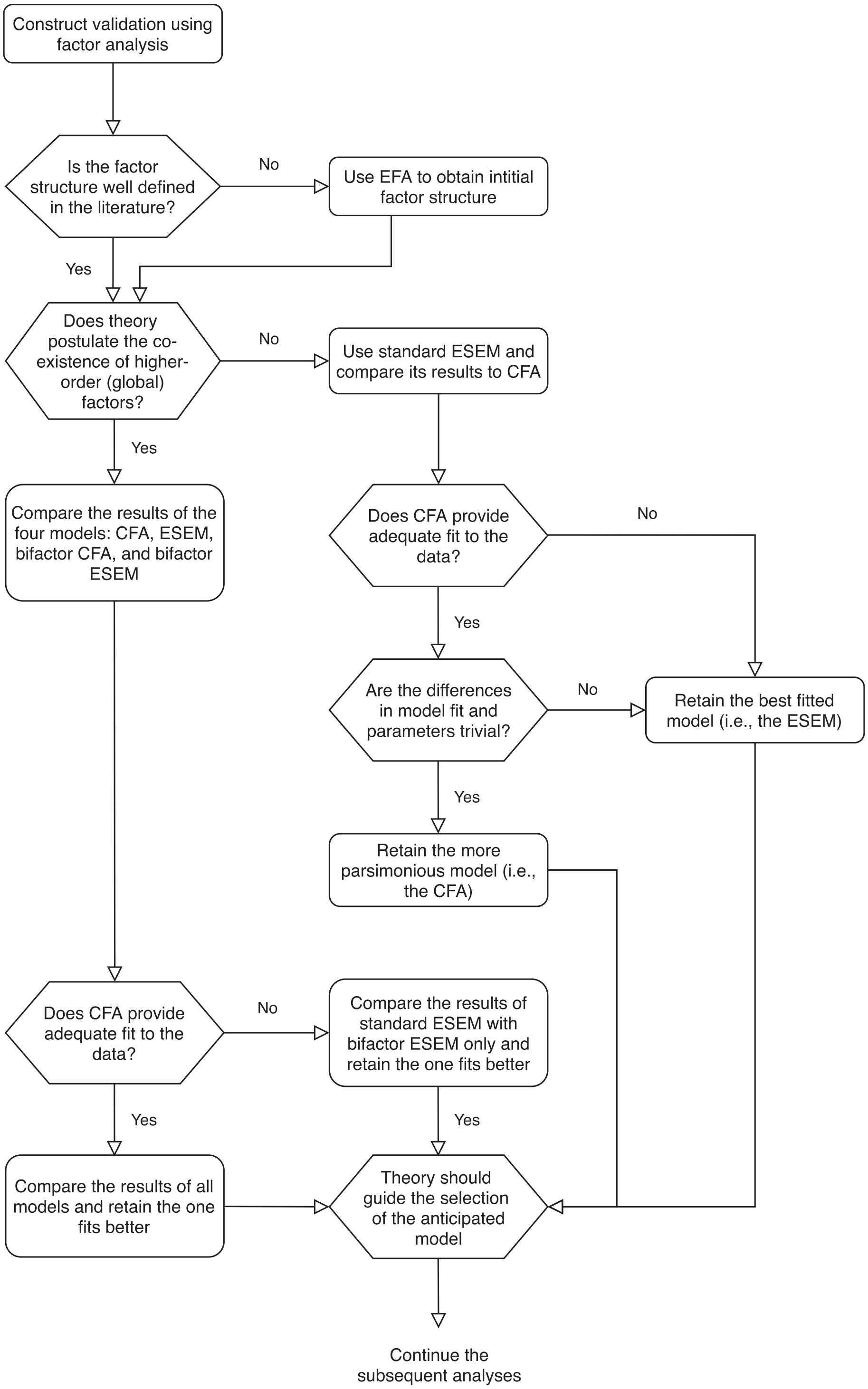

Fig. 3. A decision tree for the use of ESEM and CFA.

decreases at least by .015, and BIC and AIC values should be lower in the ESEM solution); (2) the factor correlations are reduced; (3) only small to moderate cross-loadings that are easy to justify are observed; (4) the factors, particularly the general factors, are defined by the items adequately.

In the ESEM, factor loadings on target factors should be ideally higher than .50 but values between .30 and .50 can be accepted in certain situations such as when previous psychometric research support retaining the item. Values below .30 should be considered for re-assessment in subsequent validation studies. Note that when a bifactor specification is included (such as in the bifactor ESEM) the loading is satisfactory when it meets the guidelines at least on the general factors (Morin et al., 2020). Because ESEM differs from CFA, the assessment of convergent and discriminant validity is conducted differently. Alamer and Marsh (2022) suggest that convergent validity is supported by inspecting the strength of factor loadings of the items on their target factors (either specific or general). That is, they should be significant and stronger than their loadings on other unrelated factors. In contrast, discriminant validity is established by showing that items presumed to load on the corresponding factor (either specific or general) do not cross-load substantially on unrelated factors. Discriminant validity is further supported when ESEM shows modest factor correlations compared to the inflated correlations in the CFA.

It is also important to note that ESEM requires the constructs in the assessment to be conceptually related (for example assessing three constructs of the three basic needs). In cases where constructs are conceptually unrelated, a more restrictive version of ESEM (called Set-ESEM) can be applied (Marsh et al., 2020). In Set-ESEM, items of conceptually relevant constructs are allowed to cross-load whereas items of conceptually unrelated constructs are not. In this way, the solution offers a more manageable assessment that compromises between parsimony and flexibility. Because Set-ESEM does not apply to the present study it was not included in the empirical example (see Marsh et al., 2020 for a detailed explanation). Sample size in ESEM needs also to be considered and the recommendations provided by Kline (2016) can be followed. However, given the complexity inherent in the ESEM and bifactor ESEM, researchers are encouraged to use larger sample sizes (N > 300) to avoid convergence issues. In cases where the study uses the standard ESEM with a limited number of factors (e.g., two constructs), a less stringent rule of N > 200 can be applied. Nevertheless, the concept of *statistical power* needs to be considered whenever possible. Following MacCallum et al. (1996), analysts can determine the minimum sample size needed to achieve a satisfactory value of RMSEA given the level of power. When the assumption of normality is not met, analysts are advised to use estimators that are robust against non-normality such as the maximum likelihood estimation with robust standard errors (MLR). A limitation of the current versions of Mplus (at least in the 8.3 version) is that bootstrapping procedure is not available for ESEM solutions.

Example Using the Basic Psychological Needs Theory

Self-determination theory (SDT) is a theory of human motivation that has been applied to several life domains to understand what drives individuals to pursue their goals including language learning (Ryan & Deci, 2000, 2020). SDT postulates that there are three basic psychological needs (BPN) that individuals are striving to satisfy (Ryan & Deci, 2017). When these needs are properly fulfilled, language learners are likely to engage in and volitionally practice the language in an optimal manner (Dincer et al., 2019; Noels et al., 1999, 2019; Oga-Baldwin & Nakata, 2017). Specifically, 'autonomy' refers to the feeling of having options that can be freely endorsed while learning the language. It also reflects a sense of volition in participating inside and outside the language classroom. 'Competence' refers to the feeling of effectiveness in carrying out learning tasks. It also reflects the experience of mastery while overcoming challenges in some of the language aspects. 'Relatedness' refers to learners feeling being connected with the people around them. It also reflects a sense of caring for and being cared for by others.

Research in SDT indicates that holding high levels on the BPN is associated with positive learning outcomes, such as increased autonomous motivation in learning the language (Noels et al., 1999), both inside and outside the language learning classroom (Alamer & Al Khateeb, 2021); greater vocabulary knowledge (Alamer, 2021a); improved cognitive, emotional, and behavioral engagement (Dincer et al., 2019; Oga-Baldwin et al., 2017); positive perceptions about teachers' teaching styles (McEown et al., 2014); decreased feelings of social, psychological, and proficiency anxiety (Alamer & Almulhim, 2021; Alamer & Lee, 2021; Zhou, 2016). For example, Alamer's (2021a) study used a mediational model and showed that the BPN have a direct effect on vocabulary knowledge over and above the mediation processes of effort. This specific direct effect provided unique information about the explanatory power of the basic needs for increased vocabulary learning and possibly other L2 learning outcomes. Similarly, Dincer et al., (2019) evaluated the extent to which Turkish students of English perceived autonomy-support from the teachers is linked to their BPN, which, in turn, relate to behavioral, emotional, agentic, and cognitive engagement, which, in turn, are linked collectively to L2 achievement. The study showed evidence of this chain of effects and illustrated the robustness of BPN to predict all types of engagement.

Issues related to the measures of BPN in language learning

One of the earliest attempts in the field that assessed the relevance of BPN in language learning is perhaps the study of Noels et al. (1999). Although the three basic needs have not been all considered in that study, it heralded the start of further investigations of the BPN constructs in L2 learning. Subsequent L2 articles continue to assess the relationship between the three needs usually parts of them, and different learning outcomes using different scales taken from other research domains such as general psychology. These studies apply different scales of the BPN in different L2 contexts. For example, Noels et al. (1999) used a scale developed by Pelletier and Vallerand (1996) which was designed to assess subordinate perceptions about their supervisors. In essence, Pelletier and Vallerand's (1996) scale is a 10-items measure of autonomy-supportive versus controlling environment. In a subsequent study, Noels (2001) assessed L2 students of Spanish on their self-perception of autonomy using seven items based on the

Noels et al. (1999) scale, which is, again, an adoption of Pelletier and Vallerand (1996) scale. It can be seen that three items have been dropped in this subsequent study, but no clarification was made about the removed items. Later, McEown et al. (2014) examined the BPN of 128 Canadian learners of Japanese including their perceptions of the teacher as a source of the three needs. The authors adopted Noels et al.'s (1999) scale but indicated that they used 23 items (instead of 10 or 7). That is, 13 items have been newly added in this study, but there has been hardly any attempt to test the construct validity of this modified scale, let alone the missing information of the assessment of the dimensionality of the original scale (i.e., the Noels et al., 1999 scale). Furthermore, the items of these variants of the scale are not available for the research community, so researchers have limited resources to apply the BPN instrument. Without these important details, one cannot be certain about the construct and content validity of the scale. This has led some researchers to continue borrowing scales of BPN from domains other than language learning. For instance, Alamer and Lee (2019) assessed the BPN of L2 students of English using the Ilaedi et al., (1993) scale of BPN which was basically developed for examining employee and supervisor perceptions of the employee's basic needs in the workplace domain. Even recently, studies in L2 settings continue to use scales from other domains to evaluate L2 learners' basic needs (see, for example, Dincer et al., 2019; Noels et al., 2019; Oga-Baldwin et al., 2017). Thus, the field lacks a valid and freely available scale of the BPN that is specifically designed for language learning domain.

For these reasons, Alamer (2021a) has developed a modified scale, named the BPN-L2 scale, building on the literature of BPN, and presented preliminary evidence for its validity. The BPN-L2 scale has three defined subscales representing the three basic needs of autonomy, competence, and relatedness with each having 4 items equally Alamer (2021a). evaluated the construct validity and reliability of the scale by using the higher-order CFA. The higher-order model was chosen because it permits the inclusion of second-order factor (the global BPN) which can group first-order factors (e.g., autonomy, competence, and relatedness). The analysis provided satisfactory results and the study was perhaps the first to account for the global construct of BPN in a measurement model in the field. With regards to the predictive power of the BPN, Alamer (2021a) found that BPN are linked to autonomous motivation (i.e., the quality of learners' energy in carrying out language learning activity being volitional), and directly linked to vocabulary knowledge. However, higher-order CFA models have certain limitations (as explained earlier) which limits our conceptual understanding of how this theory (reflected by its scale, BPN-L2) can methodologically represent the conceptualization of BPN which postulate two layers of constructs.

Applying bifactor ESEM to the BPN-L2 scale

Multidimensional research often reported that bifactor models (whether in ESEM or CFA) provide a better model fit to the data and capture the co-existence of the global (general) factor meaningfully (Alamer, 2021c; Morin et al., 2016, 2020; Reise et al., 2010). This is true with studies that evaluate the BPN subscales, wherein researchers observe that bifactor models are superior to the standard solutions. For example, Tóth-Király et al. (2017) has investigated the multidimensionality of the satisfaction of BPN versus frustration scale among community adults and tested its continuum structure. The researchers achieved this by contrasting different measurement models including standard CFA, ESEM, bifactor CFA and bifactor ESEM. A particular focus was on the comparison of bifactor CFA and bifactor ESEM and the findings showed that bifactor CFA solutions did not fit the data adequately in two different samples (n = 2,301) whereas the bifactor ESEM did. The researchers provided further evidence for an advantage of the bifactor ESEM by considering the test of measurement invariance, which supports the generalizability of the results across samples. Another recent validation study was conducted by Cromhout et al. (2021) who contested the adequacy of the CFA solution in establishing the psychometric properties of BPN scales. To gain more understanding of the multidimensionality of their BPN scale the authors compared the results of a bifactor CFA with those of a bifactor ESEM among South African students. Generally, their results indicate support for the bifactor ESEM solutions across different samples. In addition, the bifactor ESEM model was partially metric invariant across two of the three samples of South African students.

The BPN-L2 scale (Alamer, 2021a) has been successfully introduced to the language learning field to assess learners' basic needs while learning the L2. However, the internal structure of this scale has not been replicated using a more appropriate statistical tools such as the bifactor ESEM to capture the appearance of the global factor (the BPN). Thus, the present study aimed to replicate the factorial structure and criterion-related validity of the BPN-L2 scale in a more rigorous manner. To do so, I first compare the results of the CFA and ESEM in terms of goodness-of-fit. Second, I proceed with the solution that shows a better fit and then tests its predictive validity for autonomous motivation. Based on previous studies (e.g., Cromhout et al., 2021; Gillet et al., 2020), it is hypothesized that a bifactor ESEM would fit the data better and would be able to capture the variance in both specific and global factors meaningfully. In addition, as it is theoretically and empirically supported in the domain (Alamer & Al Khateeb, 2021; Alamer & Almulhim, 2021; Alamer & Lee, 2019; Dincer et al., 2019; Noels et al., 2019), it is postulated that in the structural bifactor ESEM model, the three specific factors as well as the global factor will positively predict autonomous motivation.

Method

The sample of the present study was 367 Saudi undergraduate ESL learners at a Saudi university. 69% of the sample was female. The institutional review board approved data collection and a convenience sample strategy was used. Students were invited through email to participate in an online questionnaire made by Google Forms. Students who did not want to participate or wanted to withdraw while participating were asked to refrain from completing the questionnaire and simply close the webpage. No missing values were observed in the data set.

Table 1
Model Fit Indices for Four Measurement Models of the BPN-L2.

Model	χ^2	*p*	*df*	SRMR	RMSEA(Low 90%/Hi 90%)	CFI	TLI
CFA	184.102	<.001	51	.10	.08 (.07/.10)	.88	.85
Bifactor CFA	76.002	<.001	42	.03	.05 (.03/.06)	.97	.95
ESEM	75.090	<.001	33	.03	.06 (.04/.08)	.96	.93
Bifactor ESEM	28.99	.22	24	.01	.02 (.00/.05)	1.00	.99

Measures

The BPN-L2 scale

The BPN-L2 scale (Alamer, 2021a) comprises 12 items measuring the three specific constructs of BPN: autonomy, competence, and relatedness (see Appendix A for the scale items). The scale was designed in a 5-point Likert-type response format. Each subscale consists of four items, constituting the three basic needs. Students were asked to indicate to what extent they agreed or disagreed with each statement. Example items are as follows: for autonomy, 'I am able to freely decide my own pace of learning in English'; for competence, 'I feel I am capable of learning English'; and for relatedness, 'My English teacher cares about my progress'.

Autonomous motivation

Students' autonomous motivation was assessed using 10 items from the SDT-L2 scale (Alamer, 2021a). The scale is based on a 5-point Likert-type format and was further validated through bifactor ESEM in another study (see Alamer, 2021b). Participants were asked to ponder the question "Why are you learning English?" and then they asked to indicate the extent to which they align with the 10 items that follows. Five items for intrinsic orientation and identified orientation were used. Example items are as follows: for intrinsic orientation, "for the satisfaction I feel when I use English"; for identified orientation, "because learning English is important for my current and future studies". Both intrinsic and identified orientations underlie the overall autonomous motivation factor.

Statistical analyses

To assess the validity and reliability of the BPN-L2 scale Jamovi (The jamovi project, 2019) and Mplus 8.1 (Muthén & Muthén, 1998) were used. First, univariate normality was inspected through skewness and kurtosis using the '+2/-2' guidelines. The reliability of the constructs was obtained by using omega ω coefficient. ω takes into account the factor loadings of the observed variables and their measurement errors, thus it suites the application of SEM. Note that omega ω tends to decrease in bifactor solutions because the variance is partitioned between the two sources of global and specific constructs. To test the goodness-of-fit, the χ^2 and its p-value were evaluated. Alternative measures were also used (Marsh et al., 2004) such as the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA) with its 90% confidence intervals (CI), and the standardised root mean square residual (SRMR). CFI and TLI values in the region of .95 indicate a good model fit, but values around .90 can be acceptable. Both RMSEA and SRMR should be equal to or lower than .07 or .05 to reflect acceptable and good model fits, respectively. MLR estimator is used to control for non-normality in the data. In the ESEM models, cross-loadings are 'targeted' to be as close to zero as possible to reflect the confirmatory approach of ESEM (Morin et al, 2020).

Results

Examining the factorial validity of the BPN-L2 scale

The results of CFA, bifactor CFA, ESEM and bifactor ESEM results are reported in Table 1. The results indicated poor fit to the data in the standard CFA model, while showing good fit in the standard ESEM. Following the decision tree showing in Fig. 1 we continue the assessment using the bifactor ESEM solution. The bifactor ESEM provided a non-significant χ^2 value as well as satisfying fit indices in CFI, TLI, RMSEA, and SRMR indicating an excellent fit even with the relatively large sample size involved in the analysis.

Although the analysis (as well as theory) suggests the use of bifactor models, it was important to inspect the factor loadings of all models to see how each solution functioned in estimating model parameters. We first contrast the CFA and ESEM as suggested by the decision tree. As shown in Table 2, the magnitudes of factor correlations seem to be larger in CFA than in ESEM. The size of correlations between autonomy and competence (CFA $r = .62$, ESEM $r = .58$), autonomy and relatedness (CFA $r = .09$, ESEM $r = .06$), and competence and relatedness (CFA $r = .05$, ESEM $r = .001$) decreased in the ESEM relative to the CFA solution. With exception to item1 on 'relatedness', the results of ESEM yield estimation of cross-loadings that are small in sizes (see Table 2). The cross-loadings were either negative or positive in direction, ranged from -.11 to .12, and were mostly non-significant. As the results have shown, the analysis of the bifactor CFA showed a good fit. However, the results of the factor loadings in the bifactor CFA were rather problematic and difficult to justify. For example, three items on 'relatedness' loaded strongly but *negatively* on their specific factor (items 2, 3, and 4). If we were to continue with this model, the only feasible way to address the issue would be through correlating measurement errors or removing unstable items. Therefore, CFA models provided either unacceptable fit or misspecification to the data while ESEM provided adequate fit indices and reduced (thus more realistic) factor correlation. For these reasons, we compare ESEM with its bifactor counterpart (i.e., the bifactor ESEM) and discard CFA models from subsequent analyses.

Table 2
Factor loadings of the four measurement models assessing the BPN-L2 Scale.

items	CFA	Bifactor CFA		ESEM			Bifactor ESEM			
	β	S-β	G-β	β	β	β	S-β	S-β	S-β	G-β
Autonomy 1	.79	.72	.51	**.94**	.18	-.01	**.71**	.01	.02	**.48**
Autonomy 2	.85	.52	.62	**.80**	.04	-.01	**.59**	.13	.01	**.55**
Autonomy 3	.54	.15	.60	**.47**	.12	.04	**.18**	-.11	.01	**.62**
Autonomy 4	.63	.27	.58	**.52**	.16	.01	**.36**	.16	.01	**.49**
ω	.80	.61		.83			.70			
Competence 1	.73	.80	.51	.12	**.59**	.02	.18	**.58**	.02	**.42**
Competence 2	.67	.40	.53	.08	**.60**	.03	.05	**.40**	.00	**.51**
Competence 3	.25	.06	.22	.00	**.21**	-.02	.03	**.10**	-.02	**.22**
Competence 4	.70	.49	.48	-.06	**.85**	-.03	-.01	**.54**	-.06	**.57**
ω	.65	.58			.81			.55		
Relatedness 1	.03	.03	.59	.20	.33	**-.01**	-.13	-.12	**-.02**	**.74**
Relatedness 2	.69	-.69	.58	.09	-.11	**.69**	.07	-.07	**.69**	**.07**
Relatedness 3	.65	-.64	.06	-.02	.11	**.65**	-.02	.04	**.64**	**.16**
Relatedness 4	.85	-.86	.15	-.06	.01	**.85**	-.03	-.01	**.85**	**.07**
ω	.67	.70	.83			.70			.72	.81

ω = omega coefficient of model-based composite reliability; target ESEM and Bifactor ESEM factor loadings are indicated in bold.

Table 3
Fit Indices for the Structural Bifactor ESEM Model Predicting Autonomous Motivation.

Model	χ^2	*p*	*df*	SRMR	RMSEA(Low 90%/Hi 90%)	CFI	TLI
Structural Bifactor ESEM	397.65	<.001	174	.04	.06 (.05/.07)	.93	.90

Table 4
Structural Bifactor ESEM Model Predicting Autonomous motivation.

Factor	β	*SE*	*p*
Autonomy	.25	.06	<.001
Competence	.50	.05	<.001
Relatedness	.18	.05	.001
The general BPN	.40	.06	<.001

The results shown in Table 1 indicate that the bifactor ESEM solution provides a substantial improvement over the standard ESEM (ΔCFI = +.04; ΔTLI = +.06; ΔRMSEA = -.03). More importantly, only the bifactor ESEM solution results in a non-significant χ^2 value. The cross-loadings in bifactor ESEM appear to be smaller than in standard ESEM (see Table 2). The bifactor ESEM solution also reveals a well-defined global factor of BPN, although weak loadings have been observed in two items on 'relatedness' (target loadings ranged from .07 to .79), but also specific factors of autonomy, competence, and relatedness have been defined adequately, although weak loadings have been observed in few items (target loadings ranged from -.02 to .85). An important observation in the bifactor ESEM is that items that did not load on their specific factors load on the global factor (see for example item 1 on 'relatedness' which loaded stronger on the global factor than on the specific factor). The inverse direction of loading is also observed; items that did not load on the global factor load on their specific factors (see for example items 2 on 'relatedness').

Testing the Criterion-related Validity of BPN-L2 Scale

To provide evidence of predictive relevance of the BPN-L2 scale a structural model based on the bifactor ESEM solution was considered. In this way, both the specific and general factors were allowed to be related to the outcome variable, autonomous motivation. Note that bifactor CFA was not included because of the issues highlighted earlier regarding its factor loadings. Before interpreting the strength of the path coefficients of the structural bifactor ESEM, I evaluated the model fit indices. As shown in Table 3, the results illustrated that the fit is acceptable after an error correlation was included between two items on autonomous motivation. As indicated in Table 4, the specific factors (i.e., autonomy, competence, and relatedness) as well as the general factor (BPN) positively and statistically predicted the outcome variable (i.e., autonomous motivation). It can be seen that the strongest predictor was competence ($\beta = .50$, $p < .001$) indicating large effect size, followed by the general BPN factor and autonomy which were moderate and modest in sizes ($\beta = .40$, $p < .001$, and $\beta = .25$ $p < .001$), respectively. Last comes 'relatedness' which was positively linked to the outcome but in a small effect size ($\beta = .18$, $p = .001$). Overall, the structural bifactor ESEM model seems to predict the outcome substantially ($R^2 = .51$) indicating that the model has a strong explanatory power.

Specific imbalance in the satisfaction of three needs

The results of the bifactor ESEM show that the global BPN factor appeared to be moderately related to autonomous motivation (β = .40, p <.001). Similarly, all specific factors of autonomy, competence and relatedness significantly predicted the outcome as well. Accordingly, these findings indicate that over and beyond learners' global levels of basic need satisfaction, the specific needs were able to explain unique variability in the outcome (i.e., autonomous motivation). That is, specific levels of imbalance in the three needs are related to further increase in autonomous motivation after controlling for the general need satisfaction. Therefore, one would argue that, beside the general need satisfaction, each specific need is equally important for predicting autonomous motivation.

Discussion

The present study aimed at presenting and examining the utility of the ESEM and bifactor ESEM that permit for the combination of EFA and CFA into a single analytical framework (Alamer & Marsh, 2022; Alamer, 2021b; Marsh et al., 2009; Morin et al., 2013, 2016). To explain how bifactor ESEM operates, I applied it to a domain-specific scale, the BPN-L2 scale (Alamer, 2021a) to assess its factorial structure in a manner that follows the theoretical underpinnings of BPN. The paper started the discussion by explaining why ESEM can be useful for validation studies compared to the CFA and provided examples of CFA studies that yielded questionable results. Then, the paper discussed the need to adopt bifactor models in situations where more than one layer of constructs is involved in the measurement model. Particularly, the paper used the recently suggested model specification within the ESEM framework, that is, the bifactor ESEM to allow for the inclusion of the global factor (Howard et al., 2018). The paper presented a decision tree for interested researchers to decide which model to support. Overall, the current study demonstrated the usefulness of the bifactor ESEM for applied linguistics and education domain by showing its application in testing the dimensionality of the BPN-L2 scale (Alamer, 2021a).

In line with previous research (Cromhout et al., 2021; Tóth-Király et al., 2017), it appeared that bifactor ESEM outperformed the bifactor CFA in the goodness-of-fit, reduced factor correlation, and well-defined factors, owning to its flexibility in the analysis (Alamer & Marsh, 2022; Marsh et al., 2020). As the results have shown, the analysis of CFA did not fit the data while the bifactor CFA did. However, the results of the factor loadings in the bifactor CFA were rather problematic and difficult to justify. This is similar to what was observed in Alamer and Lee's (2019) study where the researchers opted to remove an item on 'relatedness' to improve the fit indices (leaving the construct with three items). Another possible way to improve the fit was through considering error covariance. However, correlating multiple measurement errors in themodel or removing several items for the sake of getting better model fit is not recommended (Alamer & Marsh, 2022; Hair et al., 2019). This is because the model fit will increase but the meaning of the constructs will likely be distorted . Modification indices may even suggest the inclusion of error correlation between items of unrelated dimensions such as correlating an error between items on autonomy and competence. It is recommended not to take this step, but some researchers seem to be compelled to include such error correlations in the face of improved fit indices.

The results of the bifactor ESEM indicate that the scores on the general BPN factor and the three specific constructs can be used and interpreted, but slight caution should be taken with three items on 'relatedness' and one on 'competence' as they appear to load weakly on the general factor. This observation does not imply that the general factor is distorted but rather that most of the variance was absorbed by the specific factor, showing that these items measure the specific factor (i.e., relatedness) more than they measure the general factor (i.e., the global BPN) (see Cromhout et al., 2021 and Tóth-Király et al., 2017 for a similar observation).

Further, although it might appear that the ESEM solution is less parsimonious than the corresponding CFA (because of the number of parameters estimated), some model fit indices such as TLI and RMSEA control for parsimony (Marsh et al., 2009). Because these indices are corrected for parsimony, they can indicate that a more complex model fits less when the complexity is not needed. Hence, comparisons based on these indices address this issue. Thus, it appears that the bifactor ESEM fits the measurement of BPN-L2 scale appropriately. Based on the study findings, the theoretical framework of BPN has been supported in the bifactor ESEM model because of the inclusion of the overarching construct (i.e., the basic psychological needs), in which the items loaded properly on the general factor while also loaded on their specific factors. This preposition was not possible to be accounted for in the EFA, the standard CFA, or standard ESEM models (Alamer, 2021b; Guay et al., 2015; Morin et al., 2016).

In addition, the present study has shown that bifactor ESEM can be used to provide evidence for convergent and discriminant validity from a new empirical perspective. That is, convergent validity was achieved by showing that items were strongly loaded on their hypothesized factors, even though they were allowed to cross-load on the other factors. Discriminant validity was achieved by showing that items presumed to load on one factor did not load substantially on the others in the solution, and if cross-loaded they showed weaker loadings than on their presumed factor. Moreover, the present study was uniquely able to provide information on the predictive (criterion-related) validity of the BPN-L2 scale within the framework of bifactor ESEM. By turning the bifactor ESEM into a structural bifactor ESEM model (see Fig. 2) we were able to gain precise and high-quality results of the role of the global construct (i.e., the basic psychological needs) and the specific factors (i.e., autonomy, competence, and relatedness) in predicting scores on the outcome. Because of the flexibility of the bifactor ESEM model, the explained variance was quite high (i.e., 51% of the variance in the outcome was explained by the variables in the bifactor ESEM model). In line with BPN literature, it was found that autonomous motivation was positively and meaningfully predicted by the independent variables (Alamer, 2021a, Alamer & Al Khateeb, 2021;

Alamer & Almulhim, 2021; Alamer & Lee, 2019; Noels et al., 1999, 2019; Oga-Baldwin & Nakata, 2017). Moreover, the present study sought to assess the imbalance between the global and specific levels of psychological need satisfaction and autonomous motivation as an outcome. The findings of the present study illustrated that the global factor showed moderate effect size , and that the specific factors of autonomy, competence, and relatedness were able to have substential effects on the outcome over and beyond the variance explained by the general factor. For instance, because competence outperform other factors in the effect size, it can be postulated that specific levels of competence are needed to be met in a greater extent than the others because it was found to be linked with higher levels of autonomous motivation substantially. In contrast, the results revealed that specific levels of imbalance in satisfaction in relatedness are relatively less important for autonomous motivation. Thus, the results suggested the *imbalance* between the two layers of the psychological needs in explaining the outcome. These findings concur with those reported in Alamer and Al Khateeb, (2021), Cromhout et al. (2021), Gillet et al. (2020), Sánchez-Oliva et al. (2017), and Tóth-Király et al. (2017) and extend their findings to show how the two layers of basic psychological needs functioned among the sample of language learners.

Limitations and Conclusion

Although the present study used an advanced method to evaluate the multidimensionality of BPN in the language learning domain it has some limitations. First, the study used a relatively moderate sample size from one socio-cultural context, and therefore we would like to see replication of this validation study on other contexts using larger sample sizes. In addition, the present study relied on cross-sectional data. Future studies could benefit from using longitudinal data to account for the invariance of the measure over time. In particular, we refer the reader to Marsh et al., (2020) study of the longitudinal Set-ESEM which involved constructs from different educational domains. In addition, ESEM (and its variants) can be, currently, conducted using Mplus and partially though R using 'lavaan' package. Given that the method can be only run through these two tools one can understand why the applications of ESEM in language learning have not been fully endorsed.

ESEM in general and bifactor ESEM in particular offer methodological advancements for better assessment of the many L2 scales (Alamer & Marsh, 2022). L2 studies still suffer from the extensive reliance on the standard EFA and PCA analyses to confirm the measurement models. Nevertheless, these first-generation techniques should be only used in the early stages of scale development, such as to reduce the number of items (when this is the objective). In contrast, ESEM is a second-generation technique that provides a wealth of details and allows for deeper assessments of the factorial validity of the instruments that PCA and EFA fail to offer. For these reasons, I encourage researchers to apply ESEM for psychometric investigation and replicate previous studies that relied on EFA, PCA or even CFA. It is, however, maintained that ESEM is not a solution for ill-formulated scales but it is an analysis that overcomes limitations that are inherent in other methods. Because the analysis of the present study was conducted through Mplus I provide the syntax needed to run the bifactor ESEM and structural bifactor ESEM using target rotation (see Appendix B) with the hope that the article becomes more practical for applied researchers. There is also an online ESEM code generator that is helpful for researchers to generate the Mplus syntax for standard and bifactor ESEM (De Beer & Van Zyl, 2019). Substantively, the results inform the importance of proposing a continuum structure underlying the BPN-L2 scale on conceptual and methodological grounds. Moreover, bifactor ESEM allows researchers to assess their measures from an intuitive perspective including the evaluating the convergent, discriminant, and predictive (criterion-related) validity. The paper provided guidelines and recommendations while carrying out ESEM to help researchers apply the method in a more practical manner. Hence, I would like to see future validation studies in L2 research use ESEM and bifactor ESEM (when applicable) to obtain more accurate results of the measures.

Declaration of Competing Interest

The author declare that there is no conflict of Interest involved in the study

Acknowledgement

The author warmly thanks Alexandre Morin for his constructive feedback and invaluable contribution to the analysis and the concept of imbalance, without which the paper would not be at this level.

Appendix A
Basic Psychological Needs in Second Language (BPN-L2) Scale.

Item
Autonomy
I am able to freely decide my own pace of learning in English.
I am able to freely choose the tasks to be done while learning English.
My English teacher allows my class to choose how we approach English learning.

(*continued on next page*)

Appendix A (*continued*)

Item
My English teacher let me freely practice English in the classroom.
Competence
I feel I am capable of learning English.
I can be a successful language learner.
I am competent enough to meet the challenges and tasks posed in English learning.
I feel a sense of accomplishment in my English classes.
Relatedness
My English teacher is friendly and cordial with me.
My English teacher is very understanding (puts him/herself in other people's place) about students' problems. My classmates are willing to help and cooperate with me while learning the language.
My English teacher cares about my progress.

Appendix B

Mplus syntax (code) for running bifactor ESEM

```
VARIABLE:
    aut1 aut2 aut3 aut4 com1 com2 com3 com4
    rel1 rel2 rel3 rel4;
Model:
    AUT BY
    AUT1-AUT4
    COM1-COM4~0
    REL1-REL4~0 (*1);
    COM BY
    COM1-COM4
    AUT1-AUT4~0
    REL1-REL4~0 (*1);
    REL BY
    REL1-REL4
    AUT1-AUT4~0
    COM1-COM4~0 (*1);
Analysis:
    ESTIMATOR IS MLR;
    ROTATION=TARGET
    Output:
    STDYX;
```

Mplus syntax (code) for running the structural bifactor ESEM

```
VARIABLE:
    aut1 aut2 aut3 aut4 com1 com2 com3 com4
    rel1 rel2 rel3 rel4;
Model:
    G by
    AUT1-REL4 (*1);
    AUT BY
    AUT1-AUT4
    COM1-COM4~0
    REL1-REL4~0 (*1);
    COM BY
    COM1-COM4
    AUT1-AUT4~0
    REL1-REL4~0 (*1);
    REL BY
    REL1-REL4
    AUT1-AUT4~0
    COM1-COM4~0 (*1);
    Autom by
    ins1 ins2 ins3 ins4 ins5 ident1
    ident2 ident3 ident4 ident5;
    Autom ON G AUT COM REL;
    IDENT5 WITH IDENT3 !error correlation
Analysis:
    ROTATION=TARGET (orthogonal);
    ESTIMATOR = MLR;
output:
    STDYX;
```

References

Alamer, A. (2021a). Basic psychological needs, motivational orientations, effort, and vocabulary knowledge: A comprehensive model. *Studies in Second Language Acquisition*. 10.1017/S027226312100005X.

Alamer, A. (2021b). Construct validation of self-determination theory in second language scale: The bifactor exploratory structural equation modeling approach. *Frontiers in Psychology*. 10.3389/fpsyg.2021.732016.

Alamer, A. (2021c). Grit and language learning: construct validation of L2-Grit scale and its relation to later vocabulary knowledge. *Educational Psychology, 41*(5), 544–562. 10.1080/01443410.2020.1867076.

Alamer, A., & Al Khateeb, A. (2021). Effects of using the WhatsApp application on language learners motivation: A controlled investigation using structural equation modelling. *Computer Assisted Language Learning* http://dx.doi.org/. 10.1080/09588221.2021.1903042.

Alamer, A., & Almulhim, F. (2021). The interrelation between language anxiety and self-determined motivation; A mixed methods approach. *Frontiers in Education, 6*, Article 618655. 10.3389/feduc.2021.618655.

Alamer, A., & Lee, J. (2019). A motivational process model explaining L2 Saudi students' achievement of English. *System, 87*, Article 102133. 10.1016/j.system.2019.102133.

Alamer, A., & Lee, J. (2021). Language achievement predicts anxiety and not the other way around: A cross-lagged panel analysis approach. *Language Teaching Research*. 10.1177/13621688211033694.

Alamer, A., & Marsh, H. (2022). Exploratory structural equation modeling in second language research: An applied example using the dualistic model of passion. *Studies in Second Language Acquisition*. 10.1017/S0272263121000863.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16*, 397–438.

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.

Cromhout, A., Schutte, L., & Wissing, M.P. (2021). Factor structure and measurement invariance of the Basic Psychological Needs Scale in three South African samples: A bifactor exploratory structural equation modelling approach. *Psychological Reports*, Article 00332941211025275.

De Beer, L.T., & Van Zyl, L.E. (2019). *ESEM code generator for Mplus*. Retrieved from https://www.surveyhost.co.za/esem/ . 10.6084/m9.figshare.8320250.

Dincer, A., Yeşilyurt, S., Noels, K., & Vargas Lascano, D. (2019). Self-determination and classroom engagement of EFL learners: A mixed-methods study of the self-system Model of motivational development. *SAGE Open, 9*(2), Article 2158244019853913. 10.1177/2158244019853913.

Gillet, N., Morin, A., Huart, I., Colombat, P., & Fouquereau, E. (2020). The forest and the trees: Investigating the globality and specificity of employees' basic need satisfaction at work. *Journal of Personality Assessment, 102*(5), 702–713. 10.1080/00223891.2019.1591426.

Guay, F., Morin, S., Litalien, D., Valois, P., & Vallerand, J. (2015). Application of exploratory structural equation modeling to evaluate the academic motivation scale. *The Journal of Experimental Education, 83*(1), 51–82. 10.1080/00220973.2013.876231.

Howard, J., Gagné, M., Morin, A., & Forest, J. (2018). Using bifactor exploratory structural equation modeling to test for a continuum structure of motivation. *Journal of Management, 44*(7), 2638–2664. 10.1177/0149206316645653.

Kline, R. (2016). *Principles and practice of structural equation modeling* (6th ed.). Guilford Publications.

MacCallum, R.C., Browne, M.W., & Sugawara, H.M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130–149.

Marsh, H.W., Hau, K.T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural equation modeling, 11*(3), 320–341.

Marsh, H., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(3), 439–476. 10.1080/10705510903008220.

Marsh, H., Morin, A., Parker, P., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10*, 85–110.

Marsh, H.W., Guo, J., Dicke, T., Parker, P.D., & Craven, R.G. (2020). Confirmatory factor analysis (CFA), exploratory structural equation modeling (ESEM), and set-ESEM: optimal balance between goodness of fit and parsimony. *Multivariate Behavioral Research, 55*(1), 102–119. 10.1080/00273171.2019.1602503.

McEown, M.S., Noels, K.A., & Saumure, K.D. (2014). Students' self-determined and integrative orientations and teachers' motivational support in a Japanese as a foreign language context. *System, 45*, 227–241.

Morin, S., Arens, A., & Marsh, H. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modelling, 23*, 116–139. 10.1080/10705511.2014.961800.

Morin, S., Marsh, H., & Nagengast, B. (2013). Exploratory structural equation modeling. In G. Hancock, & R. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (pp. 395–436). Information Age.

Morin, A., Myers, N., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, exploratory structural equation modeling (ESEM) and bifactor-ESEM. In G. Tenenbaum, & R. C. Eklund (Eds.), *Handbook of Sport Psychology* (4th Edition). UK: Wiley.

Muthén, L.K., & Muthén, B.O. (1998). *Mplus User's Guide* (7thEdn). Muthén and Muthén.

Noels, K., Clément, R., & Pelletier, L. (1999). Perceptions of teachers' communicative style and students' intrinsic and extrinsic motivation. *The Modern Language Journal, 83*(1), 23–34.

Noels, K., Lascano, D., & Saumure, K. (2019). The development of self-determination across the language course: Trajectories of motivational change and the dynamic interplay of psychological needs, orientations, and engagement. *Studies in Second Language Acquisition, 41*(4), 821–851. 10.1017/S0272263118000189.

Oga-Baldwin, Q., & Nakata, Y. (2017). Engagement, gender, and motivation: A predictive model for Japanese young language learners. *System, 65*, 151–163. 10.1016/j.system.2017.01.011.

Pelletier, L.G., & Vallerand, R.J. (1996). Supervisors' beliefs and subordinates' intrinsic motivation: A behavioral confirmation analysis. *Journal of Personality and Social Psychology, 71*(2), 331.

Plonsky, L. (2014). Study quality in quantitative L2 research (1990–2010): A methodological synthesis and call for reform. *The Modern Language Journal, 98*(1), 450–470. 10.1111/j.1540-4781.2014.12058.x.

Plonsky, L. (2015). *Advancing quantitative methods in second language research*. Routledge.

Reise, S., Moore, T., & Haviland, M. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544–559. 10.1080/00223891.2010.496477.

Ryan, R.M., & Deci, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68–78.

Ryan, R.M., & Deci, E.L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness* (second edition). Guilford Publications.

Ryan, R.M., & Deci, E.L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology, 61*, Article 101860. 10.1016/j.cedpsych.2020.101860.

Sánchez-Oliva, D., Morin, A.J., Teixeira, P.J., Carraça, E.V., Palmeira, A.L., & Silva, M.N. (2017). A bifactor exploratory structural equation modeling representation of the structure of the basic psychological needs at work scale. *Journal of Vocational Behavior, 98*, 178–187.

The jamovi project. (2019). Jamovi *(1.1) [Computer software]*. https://www.jamovi.org.

Tóth-Király, I., Morin, A.J., Bőthe, B., Orosz, G., & Rigó, A. (2017). Investigating the multidimensionality of need fulfillment: A bifactor exploratory structural equation modeling representation. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(2), 267–286.