

Methods Forum

EXPLORATORY STRUCTURAL EQUATION MODELING IN SECOND LANGUAGE RESEARCH

AN APPLIED EXAMPLE USING THE DUALISTIC MODEL OF PASSION

Abdullah Alamer *

Imam Mohammad Ibn Saud Islamic University, Saudi Arabia and King Faisal University, Saudi Arabia

Herbert Marsh 

Australian Catholic University, Sydney, Australia

Abstract

This study offers methodological synergy in the examination of factorial structure in second language (L2) research. It illustrates the effectiveness and flexibility of the recently developed exploratory structural equation modeling (ESEM) method, which integrates the advantages of exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) into one complete measurement model. Two sets of data were collected using the L2 Passion Scale, which measures a dualistic model of passion. Study 1 participants were 220 L2 students. A comparison was made between the CFA and the ESEM models. The results demonstrated the superiority of the ESEM method relative to CFA in terms of better goodness-of-fit indices and realistic correlated factors. These results were replicated in another sample of 272 L2 students, providing support for the predictive validity using a structural ESEM model. Guidelines are provided and Mplus syntax files (codes) are included to help analysts apply the methods. We also make the data available publicly. Overall, this research demonstrated the usefulness of ESEM for examining the construct, discriminant, and convergent validity of L2 scales over CFA.

INTRODUCTION

Past studies that have focused on validating different psychological constructs in L2 domain have primarily used exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) to explore, develop, refine, and distinguish between latent psychological

*Corresponding author. E-mail: alamer.aaa@gmail.com

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

variables (Alamer, 2021a, 2021b, 2021c; Hancock & Schoonen, 2015; Plonsky, 2015; Woodrow, 2014). Although these two techniques generate useful information on how observed and latent variables are connected to each other, there are some limitations. Developments in factor analysis research have recently offered a technique that incorporates the utility of EFA and CFA into one analytical framework, namely exploratory structural equation modeling (ESEM) (Asparouhov & Muthén, 2009; Marsh et al., 2014; Morin et al., 2013). Thus, explaining the utility of ESEM using the constructs of the dualistic model of passion (DMP) (Vallerand, 2015; also see Horwood et al., 2021) is the focus of this article. Nevertheless, we assume that readers want to know more about the constructs of DMP to assist in contextualizing the assessment. Hence, we present an overview of the DMP. Accordingly, this research represents an essential substantive-methodological synergy that applies a cutting-edge methodology to address substantive issues with theoretical and methodological implications in L2 domain. Theoretically, this study follows from the dual passion model (Horwood et al., 2021; Marsh et al., 2014; Vallerand, 2015), positing distinct components of obsessive passion (OP) and harmonious passion (HP). Methodologically, we juxtapose traditional exploratory and confirmatory approaches to factor analysis (EFA and CFA) with ESEM that combines the most important advantages of EFA and CFA into a unified analytic approach. To make this article more practical, we are making the data publicly available, and we are including the syntax (code) necessary to run the analyses in [Appendix B](#).

EXPLORATORY FACTOR ANALYSIS AND CONFIRMATORY FACTOR ANALYSIS

As its name suggests, EFA involves exploring data searching for emerging latent variables (or factors) based on various observed variables obtained from the data (Hancock & Schoonen, 2015). As such, EFA is mostly used when researchers do not have a clear, hypothesized structure regarding the underlying factors in the data. EFA became widely used because it offered a more realistic representation of the data presumed to assess the underlying factors by allowing for the estimation of item cross-loading.

However, the empirical assessment of measurements cannot be founded solely on EFA (Marsh et al., 2014) because EFA lacks a few of the methodological advantages that CFA offers, such as computation of the goodness-of-fit indices, allowing error terms to be correlated (when theory justify this), testing the predictive power of the constructs for the outcomes, and testing the measurement invariance (equivalence) among different groups. A key specification of CFA (which is also a drawback and is discussed later) is that items that correspond to a specific factor are only allowed to load onto their hypothesized factor. Thus, cross-loadings across other presumably unrelated factors (which EFA specifies) are specified to zero (Alamer, 2021b; Tóth-Király et al., 2017). In this way, researchers use CFA when they have a hypothesized structure of the factors that exists in the data. Due to the advantages CFA offers, researchers prefer using CFA over EFA in theory-driven research because it yields a more parsimonious measurement model; this is particularly true for those who seek to reutilize previously established scales.

However, there are limitations on ignoring cross-loadings in assessing measurement models using CFA. It often results in a certain level of systematic measurement error because items do not load onto only their corresponding factor. Instead, psychological factors tend to overlap (Marsh et al., 2009, 2014; Tóth-Király et al., 2017). In many cases,

items represent not only one pure construct but also share a certain degree of commonality with other similar constructs. Hence, items can be said to be somehow imperfectly representative of their corresponding factors.

Consequently, item cross-loading should be expected and accepted. Thus, cross-loading should be assessed when conceptually related latent variables are present (Morin et al., 2016). It is quite possible that constraining cross-loading items to zero, which is a default approach in CFA, would result in an unexpressed variance that eventually inflates the correlation between the factors and distorts their definition (Marsh et al., 2009). This is because unexpressed cross-loadings on the presumed unrelated factor can only be explained through factor correlation, thus inflating the correlation size. This issue raises methodological and empirical concerns about the magnitude of the correlation between factors in the measurement model in CFA (see next section for an example).

USING EXPLORATORY STRUCTURAL EQUATION MODELING TO ASSESS CONSTRUCT VALIDITY

ESEM has been developed as an improved method that overcomes the aforementioned EFA and CFA limitations (Asparouhov & Muthén, 2009; Horwood et al., 2021; Marsh et al., 2009; Morin et al., 2013). In essence, ESEM simultaneously permits cross-loadings among items and factors (which is an EFA-derived property), and computes goodness-of-fit indices (which is a CFA-derived property). Thus, ESEM integrates the advantages of both EFA and CFA into one complete model (Marsh et al., 2014). In fact, one could imagine ESEM as an EFA being executed through SEM. Hence, ESEM takes on all advanced features that SEM can offer. However, to maintain the confirmatory application of ESEM, a ‘target rotation’ may be used (Morin et al., 2016). In target rotation, cross-loadings are ‘targeted’, but not forced, to be close to zero. Figures 1 and 2 comparatively illustrate the differences between the CFA and ESEM models visually.

It is evident that the CFA model is more restrictive than ESEM because cross-loadings are constrained at zero. This issue has resulted in several cases of less-than-optimal support for the constructs validity in CFA that had been established in EFA. For example, Marsh et al. (2009) illustrated that standard CFA models often fail to meet the new, even more demanding, cutoff guidelines in establishing the measurement model. Thus, L2 researchers appear to be forced to apply *ex post facto* modifications to get a better fit that,

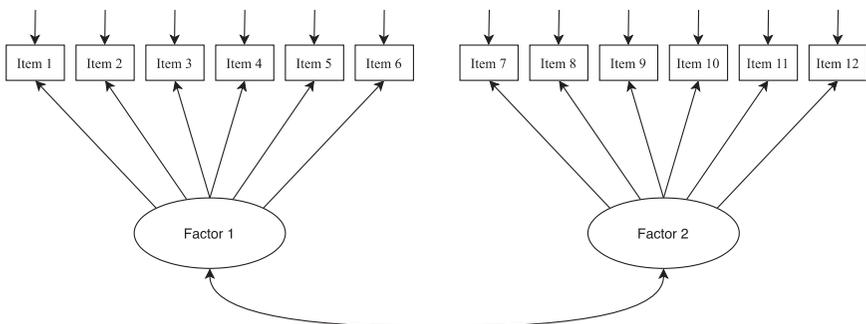


FIGURE 1. A simplified representation of the CFA model.

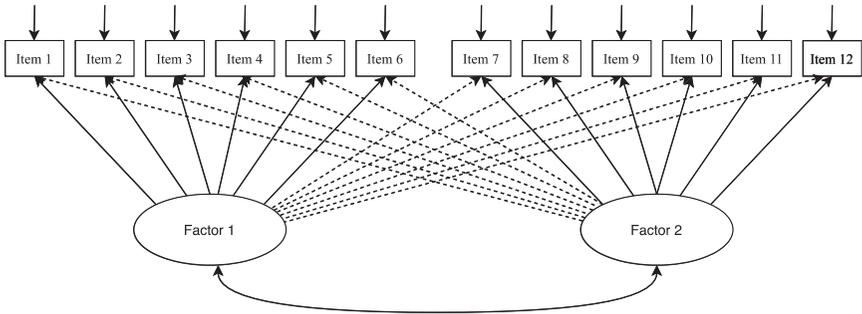


FIGURE 2. A simplified representation of the ESEM model.

in turn, may impact the measurement theory. Consequently, researchers stress that even small cross-loadings (such as .10 or .15) should be considered; otherwise, path coefficients and factor correlations could potentially be inflated and biased, thus affecting the goodness-of-fit (Guay et al., 2015). Alamer and Lee (2019) offered an example from L2 literature that points to the issue of CFA models as the researchers examined the measurement model of Self-Determination Theory. It was anticipated that five motivational factors—namely intrinsic, identified, introjected, external regulations, and amotivation—would emerge in the CFA; however, the model fit indices indicated a poor fit to the data (i.e., $\chi^2 = 705.11$, RMR = .11, RMSEA = .09, CFI = .91, NFI = .88), leading the researchers to reformulate the measurement model. A similar concern has been recently raised in the L2 domain by Alamer (2021b). The researcher provided examples from L2 literature that assessed a scale of Self-Determination Theory in language learning using the standard CFA. Alamer (2021b) indicated that these studies tended to present inflated factor correlations that are contrary to theory. For instance, a previous study (Ardasheva et al., 2012) reported that the correlation between intrinsic regulation and external regulation in the CFA was positive and substantial $r = .87$ which has been increased substantially from $r = .55$ in the EFA (i.e., $\Delta r = +.32$). These issues, however, do not imply that researchers should refrain from using CFA and rely solely on EFA. Rather, the combined ESEM framework can be a synergic methodological approach to overcome the previously mentioned CFA and EFA issues. A similar concern about relying on CFA and ignoring ESEM in assessing the psychometric properties of the constructs has been also raised in Alamer (in press).

WHEN TO USE ESEM?

Although its name may suggest an exploration procedure, ESEM is mainly used to confirm *a priori* defined factor structure. Therefore, it might be more appropriate for newly developed scales to use EFA, especially when the researcher's objective is to search for the relationships among the variables to determine the possible number of factors. ESEM should be, then, used to confirm the factor structure that EFA suggested. In cases in which the factor structure is already defined and tested in previous research, the researcher can skip the EFA and conduct the ESEM analysis to assess the model's fit to the data. ESEM should

then be compared to CFA. If the ESEM and CFA models are similar in fit and parameter estimates, the researcher should prefer the more parsimonious CFA model. Nonetheless, not all measurement models should apply the standard ESEM. In some cases, the theory may postulate the coexistence of the global factor(s). In such cases, the researcher should include the bifactor CFA and bifactor ESEM in the comparison (see Alamer, 2021b, *in press* for examples in the L2 domain). In addition, when the design of the study or intended purpose dictates that one set of factors should not cross-load on another, Set-ESEM should be used (Marsh et al., 2020). For example, when the study includes longitudinal data, the researcher can keep the constructs separate for each wave (see Marsh et al., 2020 for details). The tree decision shown in Figure 3 illustrates the process of selecting the appropriate measurement model.

To the best of our knowledge, only three empirical studies in the applied linguistics and language learning domain have employed ESEM to assess multidimensionality (Alamer, 2021b, *in press*; Alamer & Almulhim, 2021). These studies showed that ESEM and bifactor ESEM were successful and outperform their CFA counterparts. The present study demonstrates the power of the ESEM method and applies it to the conceptualization of the DMP in two samples of language learners.

THE DUALISTIC MODEL OF PASSION AND THE L2 PASSION SCALE

Persons are more likely to learn a second language successfully when they are passionate about the undertaking (Vallerand, 2015). Research on the concept of passion has been conducted primarily within the DMP framework (Horwood et al., 2021; Marsh et al., 2013; Vallerand, 2015, 2016; Zhao et al., 2015; but also see Pollack et al., 2020). DMP postulates the existence of two distinct types of passion, namely harmonious passion (HP) and obsessive passion (OP), each of which accounts for one of two paths, leading to either positive or negative consequences. For more than a decade, the DMP has been used to study passion across various life domains such as sports, marketing, teaching, work, social activities, and leisure (see Vallerand, 2015, for a review). In addition, several studies have demonstrated empirical support for the DMP and the Passion Scale—a scale that is developed to assess the two types of passion (see Horwood et al., 2021; Vallerand, 2015, 2016, for a review). However, there have been very few attempts to examine the constructs of DMP in the field of language learning. Among these is the study of the role of HP and OP among Chinese L2 students (Chen et al., 2021). The study involved 260 Chinese high school students learning English as an L2. The researchers found that HP better predicted mastery goals and L2 flow than did OP. In turn, L2 flow experience mediated the effect of HP and OP on willingness to communicate.

In essence, DMP postulate that passion is “a strong inclination toward a specific object, activity, concept, or person that one loves (or at least strongly likes) and highly values, that is part of [one’s] identity, and that leads one to invest time and energy in the activity (or person) on a regular basis” (Vallerand, 2015, p. 331). Passion is not only directed toward the activity (e.g., learning a language), but it goes beyond that to also include other aspects related to the activity such as a person (e.g., a native speaker) or a specific object (e.g., the targeted language). In this manner, passion is inherently motivational in nature. Indeed, there is a desire, a goal, an inclination, an impulse, and a drive toward an object that is perceived to be

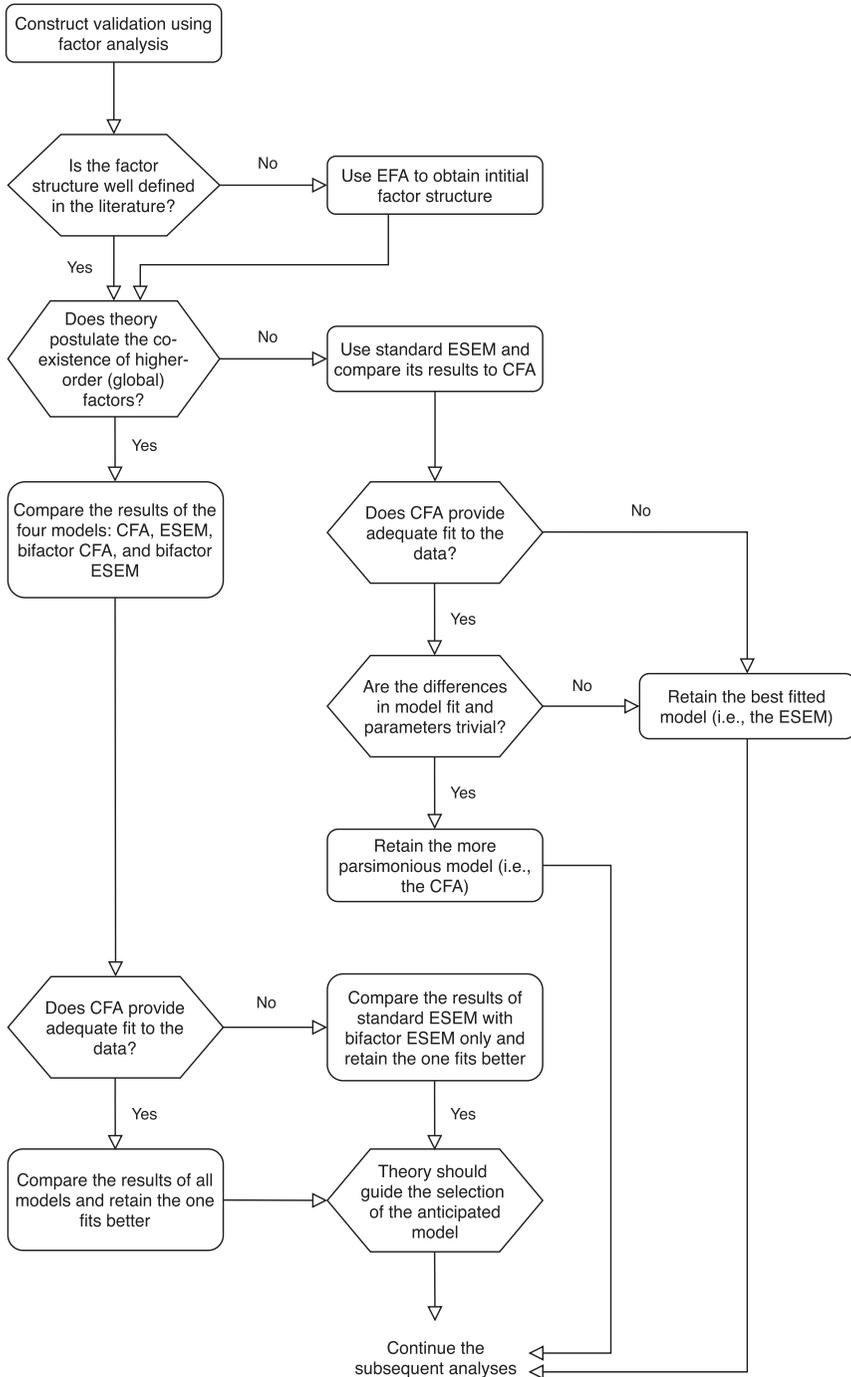


FIGURE 3. Decision tree for the use of the appropriate measurement model.

desirable. However, passion can also result in negative consequences, such as losing personal control, inflexible persistence, emotional suffering, and the erosion of well-being. Therefore, the DMP postulates the existence of the two previously mentioned forms of passion, HP and OP. The difference between these two types of passion rests primarily on the internalization of the passionate activity (Vallerand, 2015). As per self-determination theory, internalization is a process that reflects the extent to which socially sanctioned values and behaviors are perceived autonomous (Ryan & Deci, 2017).

Thus, HP results from an autonomous internalization of the tasks that are integrated into one's identity. According to self-determination theory, autonomous internalization is a term that applies to individuals who have decided, of their own volition, that an activity is valuable and important (Ryan & Deci, 2000). Such internalization originates from within the self and integrates well with one's identity. Thus, HP is likely to produce a motivational drive that is related to engagement in the activity. It also fosters feelings of autonomy and promotes the meaningful pursuit of the activity (Vallerand, 2016). With regard to HP, the activity is generally integrated with one's self-structure, and is in harmony with the other elements of the individual's life. Therefore, people with high levels of HP should wholeheartedly engage in the passionate activity with a nondefensive attitude (Brown & Ryan, 2003). In so doing, individuals with high levels of HP will be able to immerse themselves in the task at hand and experience positive outcomes both during the execution of the task (e.g., expending greater effort) and after the task's completion (e.g., satisfaction and willingness to continue involvement in the task) (Vallerand, 2016).

In contrast to HP, OP is generated from the controlled internalization of a task (Vallerand et al., 2003). That is, the task is not fully internalized in the self and there may be some contingencies attached to it (Brown & Ryan, 2003; Ryan & Deci, 2017). Such controlled internalization originates from an intrapersonal and/or interpersonal urge because particular contingencies are attached to the activity (e.g. social pressure, self-esteem, or an uncontrollable excitement derived from being involved in the activity; Horwood et al., 2021; Marsh et al., 2013; Zhao et al., 2015). Therefore, although OP entails passion for a particular activity, individuals with an OP may experience an uncontrollable urge to partake in an activity. Accordingly, individuals with high OP levels may risk experiencing conflicts and losing the flow (Zhao et al., 2015).

THE CONSTRUCT VALIDITY OF THE L2 PASSION SCALE

Many empirical investigations that have been conducted across various life domains have supported the usefulness of the DMP (see Vallerand, 2015 for a review). Furthermore, studies that have examined the psychometric properties of its scale (the Passion Scale) have demonstrated strong support for its reliability and validity. EFA, CFA, and even ESEM support the two-factor structure of the Passion Scale as well as its invariance across genders, languages, and different types of activities (Fernet et al., 2014; Horwood et al., 2021; Marsh et al., 2013; Vallerand et al., 2003; Verner-Filion et al., 2017; Zhao et al., 2015). For example, using the Passion Scale, Tóth-Király et al. (2017) explored the ESEM method's applicability to assessing Hungarian adults' passion for using social media and playing on smartphones and with online games. The study compared CFA and ESEM results, confirming ESEM's superiority in terms of goodness-of-fit indices and the definition of the factors. The researchers found that the correlation between HP and OP in ESEM was

relatively modest but meaningful, representing the communalities between the factors more accurately than CFA. More recently, Chen et al. (2021) have tested the relationship between HP and OP, and L2 positive experience when carrying out L2 activity. The study examined a CFA model that included both HP and OP. Their results have shown that the initial CFA model was unacceptable (e.g., CFI = .90, TLI = .87, and RMSEA = .14). To improve the model's fit, the researchers have correlated all uniquenesses (error terms) in the model, with exception of two items. After correlating multiple error terms, their model appeared to be improved (e.g., CFI = .97, TLI = .95, and RMSEA = .08). However, it is still not clear whether ESEM can be a useful tool to overcome the need to apply multiple post hoc modifications (e.g., correlating the error terms) in the measurement model to achieve an acceptable fit. Although the scale used in Chen et al. (2021) is similar to the present study's scale, the main difference is related to the fact that the Chen et al. (2021) scale has been translated to Chinese whereas the present study maintained the English version (i.e., it has not been translated to participants' L1) but the items have been adjusted to suit language learning passion. Thus, the present study builds directly on the Passion Scale (Vallerand et al., 2003) and introduced the language-specific scale (L2 Passion Scale) by adjusting the wording of the items in the original scale to suit the language learning setting.

Further, empirical research within the DMP framework has underscored the positive relationship between the two types of passion and *passion criteria*. In essence, the passion criteria were based on the definition of passion (Vallerand et al., 2003). They allow interested researchers to examine the extent to which an individual loves/likes and values an activity. In addition, passion criteria also assess the time and energy that is invested into the activity and whether the individual perceives the activity to be a part of him/herself and considers it his/her passion. Therefore, research studies have consistently reported that both HP and OP are positively associated with passion criteria, thus supporting the predictive validity of the two types of passion (Horwood et al., 2021; Guay et al., 2015). Nonetheless, past research has shown HP to be a substantially better predictor of passion criteria than OP (Zhao et al., 2015).

THE PRESENT STUDY

The present study aimed to apply the ESEM method to L2 research, using the DMP framework. To achieve this, we adopted the two-studies design. In Study 1, we psychometrically analyzed the L2 Passion Scale among a sample of Saudi EFL students. We first conducted an EFA on the data that we obtained. After establishing the suggested factorial structure from the EFA, we applied both CFA and ESEM and compared their results to obtain the best-fitting measurement model. We then replicated the results of Study 1 in Study 2 with another sample of Saudi EFL students, and we extended the application of ESEM by assessing the predictive validity of the L2 Passion Scale. We based the predictive validity on the best-fitting model obtained from Studies 1 and 2, and the extent to which HP and OP can predict L2 passion criteria scores (Vallerand et al., 2003) using an ESEM structural model. Kline (2016) indicated that 200 cases are the median sample size in uncomplex models, but cases > 200 are generally desirable. The samples for Studies 1 and 2 meet this guideline.

In making this study more practical for the readers, we make the data available publicly to rerun the data using the methods explained in the article. In addition, we provide the

Mplus syntaxes (codes) required to run CFA, ESEM, and structural ESEM models (see Appendix B).

RESEARCH QUESTIONS

RQ1: What factor structure model (i.e., CFA or ESEM) best represents the L2 Passion Scale?

RQ2: Does ESEM provide support for the convergent, discriminant, and predictive validity of the L2 Passion Scale?

STUDY 1

PARTICIPANTS AND PROCEDURE

A total of 220 Saudi EFL undergraduate students enrolled in the Department of English at a Saudi public university received email invitations to participate in this study. The participants' ages ranged from 20 to 26 years old, with a mean age of 21.3 years ($SD = .46$). The sample was 39% male, and we only contacted students who had completed their foundation year and entered the department's program. Students in the foundation year had to pass English subjects designed for A2 English level—based on the Common European Framework of Reference for Languages (CEFR).

The institutional review board approved data collection. An e-mail request to complete an online Google Forms questionnaire was sent to the students. An inspection of the responses revealed that none of the participants had a lower completion rate of 90%; therefore, no data were deleted. Missing data were handled by implementing the full information maximum likelihood (FIML) estimation. There were only five instances of missing data for only two of the variables, so missing data was not a serious issue.

MEASURES

We examined the reliability of the measures through McDonald's composite reliability (ω). Composite reliability ω (also known as CR) differs from Cronbach's alpha (α) because the former takes into account the magnitude of association between variables as well as their error measurements, therefore assessing the extent to which different variables represent the overall construct (McDonald, 1970). Values equal to .70 or above indicate acceptable internal consistency of reliability. However, compared to CFA, the values of ω tend to decrease in ESEM, and thus researchers should expect and accept this observation as long as other model parameters are comparable. To evaluate the quality of the models, we consider the chi-square statistic χ^2 and its p -value. However, χ^2 often yields significant results in large samples. Thus, we evaluated goodness-of-fit for all models based on traditional indices that are reasonably independent of sample size (RMSEA = root mean square error of approximation with its 90% confidence interval, CFI = confirmatory fit index, TLI = Tucker–Lewis index, and SRMR = standardized root mean square residual). 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 acceptable and good model fits, respectively (Hair et al., 2019; Marsh et al., 1988). For comparing nested models,

Cheung and Rensvold (2002) suggested that if the decrease in fit for the more parsimonious model is less than .01 for incremental fit indices such as the CFI, there is reasonable support for the more parsimonious model. For indices that incorporate a penalty for lack of parsimony, such as the RMSEA and the TLI, it is also possible for a more restrictive model to result in a better fit than would a less restrictive model. For present purposes, to facilitate communication, we present primarily TLIs in the written summaries of the results. In addition, however, we present the chi-square, degrees-of-freedom, RMSEA, CFI, and TLI in the corresponding tables. Nevertheless, these cutoff values for these indices constitute only rough descriptive guidelines rather than “golden rules” (Marsh et al., 2004).

The L2 Passion Scale

This L2-specific scale was developed based on the original Passion Scale that Vallerand et al. (2003) developed and Marsh et al. (2013) subsequently refined. *The L2 Passion Scale* items have been slightly modified to suit the L2 context. The L2 Passion Scale uses two six-item subscales that assess HP and OP, similar to the original scale. The students responded to all the statements based on a 5-point Likert scale, ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*). An example of an item that assesses HP in the context of L2 is as follows: “Learning English is in harmony with the other activities in my life.” Similarly, an example of an item that assesses OP for L2 learning is as follows: “I have difficulties controlling my urge to learn English” (see [Appendix A](#) for scale items). The mean and standard deviation were $M = 4.28$, $SD = .53$.

RESULTS

NORMALITY AND OUTLIERS

It is important to check whether the data are normally distributed. To test univariate normality, we examined the skewness and kurtosis of the data distribution using the A “+2/-2” guideline (Collier, 2020). Our data did not violate these guidelines. We tested multivariate outliers using the Mahalanobis D^2 measure for all variables in the data set, which instructs that any data point that has a D^2 value that deviates substantially from the dataset’s D^2 values (i.e., $p < .001$) should be removed (Hair et al., 2019). The results showed that five cases exceeded the acceptable threshold; therefore, they were excluded from the analyses.

THE MEASUREMENT MODEL OF THE L2 PASSION SCALE: CFA AND ESEM

Based on the results obtained from the EFA (see [Appendix C](#)), the next step was to confirm the hypothesized factor structure of the L2 Passion Scale and assess its construct validity by comparing the results of CFA with those of ESEM. All measurement models were estimated using the maximum likelihood estimation with robust standard errors in Mplus 8.1 (Muthén & Muthén, 2014). In the CFA model, the items were specified to load onto their *a priori* construct (i.e., HP or OP). However, in the ESEM model, the items were allowed to be freely estimated and they cross-loaded onto HP and OP (see [Figures 1 and 2](#)). A Geomin rotation was used in the ESEM.

TABLE 1. CFA and ESEM model fit indices for the L2 Passion Scale in study 1

Model	χ^2	<i>p</i>	<i>df</i>	SRMR	RMSEA	RMSEA Low 90%	RMSEA High 90%	CFI	TLI
CFA	287.790	<.001	53	.13	.15	.13	.16	.62	.53
ESEM	180.163	<.001	43	.07	.12	.10	.14	.78	.66
CFA ^a	111.873	<.001	52	.08	.07	.06	.09	.90	.88
ESEM ^a	76.519	<.001	42	.05	.06	.04	.08	.95	.91

^aCorrelated uniqueness between OP1 and OP2.

The results presented in Table 1 indicate that both CFA and ESEM solutions generated an unsatisfactory model fit with the data. Furthermore, an inspection of the modification indices using Mplus 8.1 suggested that the inclusion of covariance between the uniqueness of OP1 “If I could, I would only do my English learning tasks” and OP2 “I have almost an obsessive feeling for learning English” would significantly improve model fit (items correlated strongly, $r = .71, p < .001$). This covariance was considered in the two solutions, resulting in an improved model fit in both CFA and ESEM (see Table 1 under CFA and ESEM). This inclusion was also justified considering that the two items might share a certain similarity and an inspection of the wording of the items suggests that the inclusion of the correlated uniquenesses is reasonable. Although both solutions provided the expected L2 Passion Scale factor structure, only the ESEM provided an adequate fit to the data (see Table 1 for details). Furthermore, the differences between the fit values for CFA and ESEM were above the typical criterion (e.g., difference in CFI > .015) used to test the difference between nested models. These results support the ESEM solution’s

TABLE 2. Factor loadings of CFA and ESEM based on the most adequate solutions

Items	CFA (uniq)	ESEM		(uniq)
		HP	OP	
HP items				
HP1	.44 (.81)	.47	-.07	(.79)
HP2	.69 (.57)	.66	.07	(.53)
HP3	.69 (.53)	.65	.09	(.55)
HP4	.71 (.49)	.72	.01	(.47)
HP5	.58 (.66)	.63	-.10	(.62)
HP6	.48 (.77)	.47	.02	(.77)
ω	.77			
OP items				
OP1	.23 (.95)	-.21	.33	(.88)
OP2	.33 (.89)	-.06	.36	(.87)
OP3	.55 (.70)	.29	.45	(.65)
OP4	.45 (.79)	-.07	.49	(.77)
OP5	.76 (.41)	.12	.72	(.42)
OP6	.81 (.35)	.00	.82	(.31)
ω	.71	.64	.57	
Correlation: HP and OP	.34		.23	
correlated (uniq): OP1 and OP2	.70		.69	

Note: Nonsignificant parameters ($p \geq 0.05$) are italicized; (uniq) = item uniqueness; HP = harmonious passion; OP = obsessive passion; ω = omega coefficient of model-based composite reliability.

superiority to CFA for improved (and acceptable) goodness of fit. This, in turn, supports the L2 Passion Scale's construct validity within the ESEM framework.

We present standardized factor loadings in [Table 2](#). We note that the items loaded on their presumed factors for both the CFA and the ESEM solutions. Most items loaded strongly on their respective factors in the ESEM; however, there was one instance of positive cross-loading (i.e., OP3 loaded significantly on HP and OP, though weakly on HP, $\beta = .29$) that might appear puzzling at first, although it did not impair the definition of the two factors. Indeed, this parameter might offer useful information on how OP and HP share some degree of similarity. Hence an item (or items) on one passion factor could be empirically loaded weakly on another (Tóth-Király et al., 2017; Vallerand et al., 2003; Zhao et al., 2015). Although OP1 and OP2 factor loadings were modest, we retained the items in the analysis because they are part of the original scale. Furthermore, removing any of these items will significantly decrease reliability estimates and potentially distort the meaning of the construct (Hair et al., 2019).

We assessed convergent validity by inspecting the strength of items loading on their respective factors in the ESEM solution (Guay et al., 2015). The results presented in [Table 2](#) support these relationships because the items were strongly loaded on their hypothesized factors, even though they were allowed to cross-load on the other factor. Validity was also supported by the small correlation between the HP and OP factors in ESEM ($r = .23, p = .02$). Discriminant validity was supported in that items presumed to load on one factor did not load on the other in the ESEM solution, and if there were cross-loadings they were weaker than the target loadings. Another important result that was obtained from the ESEM is related to the factor correlation between HP and OP, which was reduced from ($r = .34, p < .001$) in the CFA solution to ($r = .23, p = .02$) in the ESEM solution. This decrease can be attributed to the variance expressed in the cross-loadings between the factors that were allowed in ESEM and not in CFA. This decreased correlation provides a more accurate representation of the relation between HP and OP because the association was not inflated by ignoring the cross-loadings (Marsh et al., 2014).

STUDY 2

PARTICIPANTS AND PROCEDURE

A new sample consisting of 272 Saudi EFL undergraduate students was drawn from the same institution as the Study 1 sample and invited to participate voluntarily. Their ages ranged from 18 to 22 years old, with a mean age of 19.2 years ($SD = .26$), and the sample was 62% male. In Study 2, we contacted only students studying in the foundation year of the department of English. In this way, the researchers ensured that the participants are independent of Study 1. Data collection was approved by the institutional review board. This data collection occurred 5 weeks after we conducted Study 1. An e-mail was sent to the students inviting them to participate in the study, which required them to complete an online Google Forms questionnaire. An inspection of the responses revealed that none of the participants had a completion rate that was lower than 90%; therefore, their data were retained in the analyses from the analyses. Thus, missing data was not a serious issue and appropriately handled with FIML.

MEASURES

The L2 Passion Scale

In Study 2, we used the same measure as in Study 1.

The L2 passion criteria

This measure was adapted from the Passion Scale (Vallerand, 2015) and five items. The scale examines whether individuals are passionate about a given activity (the scale has been adjusted to suit the language learning setting). A 5-point Likert scale ranging from 1 (*Strongly disagree*) to 5 (*Strongly agree*) was used. An example item is “I like my English learning” (see Appendix A for scale items). The mean and standard deviation were $M = 4.30$, $SD = .61$.

RESULTS OF STUDY 2

NORMALITY AND OUTLIERS

The normality and outlier tests were conducted in Study 2. Our data did not violate univariate normality; however, a multivariate outlier analysis using the Mahalanobis D^2 measure for all variables involved showed that nine cases exceeded the acceptable threshold (i.e., their D^2 values deviated substantially from the rest at $p < .001$). Therefore, these cases were excluded from further analyses.

REPLICATING THE MEASUREMENT MODEL OF THE L2 PASSION SCALE USING ESEM

Study 1 demonstrated the ESEM’s solution superiority to the CFA solution. In Study 2 we sought to replicate these results by comparing the CFA and the ESEM models. We then examined predictive validity based on the more appropriate model. The results of the CFA and the ESEM goodness-of-fit indices are presented in Table 3, and the models’ factor loadings are provided in Table 4. Similar to what was observed in Study 1, the modification indices generated using Mplus software suggested the inclusion of a covariance between the uniqueness of OP1 and OP2 in both CFA and ESEM, which were considered—replicating the results from Study 1. We also note that Chen et al. (2021) also reported the need to include this correlated covariance between these two items in their L2 passion study. Overall, the results of Study 2 replicated the advantages of ESEM over CFA, as indicated by the lower chi-square (χ^2) statistic, the better

TABLE 3. CFA and ESEM model fit indices for the L2 Passion Scale in Study 2

Model	χ^2	p	df	SRMR	RMSEA	RMSEA Low 90%	RMSEA High 90%	CFI	TLI
CFA	346.149	<.001	53	.13	.15	.13	.16	.68	.60
ESEM	233.891	<.001	43	.07	.12	.11	.15	.76	.64
CFA ^a	143.210	<.001	52	.08	.08	.07	.10	.90	.87
ESEM ^a	96.134	<.001	42	.04	.07	.05	.09	.93	.90

^aCorrelated uniqueness between OP1 and OP2.

TABLE 4. CFA and ESEM factor loadings based on the most adequate solutions

Items	CFA (uniq)	ESEM		
		HP	OP	(uniq)
<i>HP items</i>				
HP1	.47 (.78)	.48	-.06	(.77)
HP2	.67 (.55)	.66	.07	(.54)
HP3	.66 (.56)	.63	.10	(.57)
HP4	.72 (.49)	.72	.00	(.47)
HP5	.64 (.59)	.76	-.07	(.57)
HP6	.50 (.75)	.47	.07	(.72)
ω	.78			
<i>OP items</i>				
OP1	.24 (.94)	-.21	.31	(.88)
OP2	.35 (.88)	-.08	.37	(.86)
OP3	.57 (.67)	.27	.51	(.63)
OP4	.46 (.79)	-.08	.48	(.77)
OP5	.73 (.46)	.12	.70	(.47)
OP6	.83 (.31)	.00	.84	(.29)
ω	.71	.65	.59	
Correlation: HP and OP	.28		.17	
correlated (uniq): OP1 and OP2	.70		.69	

None: Nonsignificant parameters ($p \geq 0.05$) are italicized; (uniq) = item uniqueness; HP = harmonious passion; OP = obsessive passion; ω = omega coefficient of model-based composite reliability.

goodness-of-fit measures (see Table 3), the meaningful cross-loading parameters, and less inflated factor correlation (see Table 4). These findings provide further evidence of the construct validity of the L2 Passion Scale’s measurement model as well as its convergent and discriminant validity.

TEST OF MEASUREMENT INVARIANCE

Measurement invariance between the samples of Study 1 and Study 2 was conducted to substantiate psychometric invariance of the solution across the two samples. A total of 492 cases (220 for Study 1 and 272 for Study 2) were considered in a multigroup comparison. We, first, run CFA and ESEM models with the merged samples to see if the models fit the data. The CFA solution, nonetheless, failed to converge while ESEM provided acceptable fit to the data (i.e., $\chi^2 = 148.223$, $df = 42$, $p < .001$, CFI = .93, TLI = .89, RMSEA = .07, RMSEA 90% CI [.06, .09], SRMR = .04). The configural, metric, and scalar models of the ESEM fit the data (e.g., CFI/TLI $\geq .90$, RMSEA $\leq .07$ and

TABLE 5. Test of measurement invariance between the samples of Study 1 and Study 2

Model comparison	$\Delta\chi^2$	p	df	Δ RMSEA	Δ CFI	Δ TLI
Configural \times Metric	1.871	.999	20	.01	.012	.001
Configural \times Scalar	2.458	.985	30	.01	.011	.01

SRMR $\leq .07$) (for a greater explanation about configural, metric, and scalar models, see Collier, 2020 and Kline, 2016; also see Marsh et al., 2020 in relation to invariance tests with ESEM). Thus, we turn to the most important part of the assessment, the comparison results. The first measure that demonstrates psychometric invariance is the p -value of the $\Delta\chi^2$ across the models. The results, as shown in Table 5, suggest that all specified equality constraints are tenable. In addition to the $\Delta\chi^2$, measures of Δ RMSEA, Δ CFI, and Δ TLI are checked. The changes in these values across the models were trivial and support measurement invariance.

ASSESSING THE PREDICTIVE VALIDITY OF THE L2 PASSION SCALE USING THE ESEM METHOD

The present study extended ESEM's effectiveness by assessing the L2 Passion Scale's predictive validity in predicting passion criteria scores (Vallerand, 2015). We present (Figure 4) a simplified representation of this hypothesized structural model. This ESEM structural model was competed against a CFA structural model (see Figure 5). The complete structural ESEM-based model outperformed the CFA model and resulted in a good fit to the data. Thus, we continue analyzing this model only (see Table 6). Note that the factor loadings have been already reported in Table 4 and we now only focus on the results of prediction. The results presented in Table 7 show the robustness of the HP and OP to predict the L2 passion criteria. This ESEM structural model explained a large amount of the variance (69%) in the L2 passion criteria for both the HP and the OP factors. Although both HP and OP significantly predicted the L2 passion criteria, it was found that HP outperformed OP in predicting the outcome. Overall, the ESEM structural model results suggest that both the HP and the OP factors reflect the conceptualization of passion (albeit differently in terms of magnitude), as represented in the DMP literature (Vallerand, 2015).

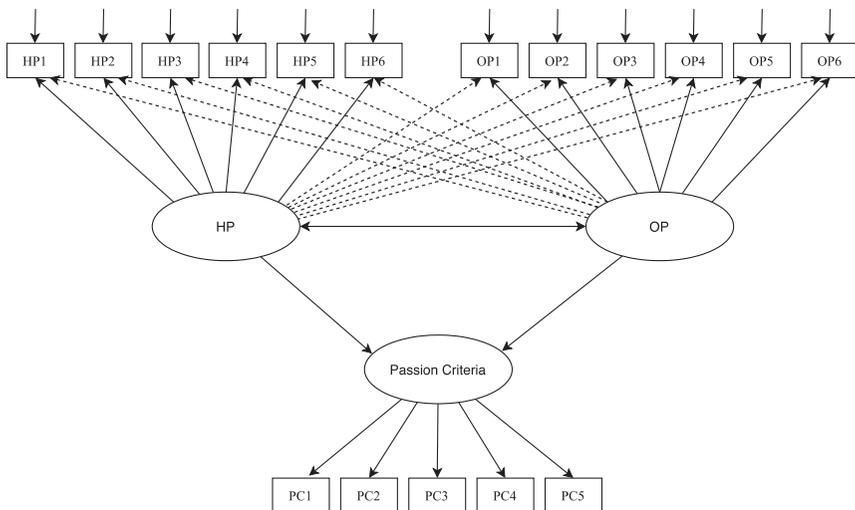


FIGURE 4. A complete structural model based on the ESEM solution.

Note: HP = harmonious passion; OP = obsessive passion; PC = passion criteria.

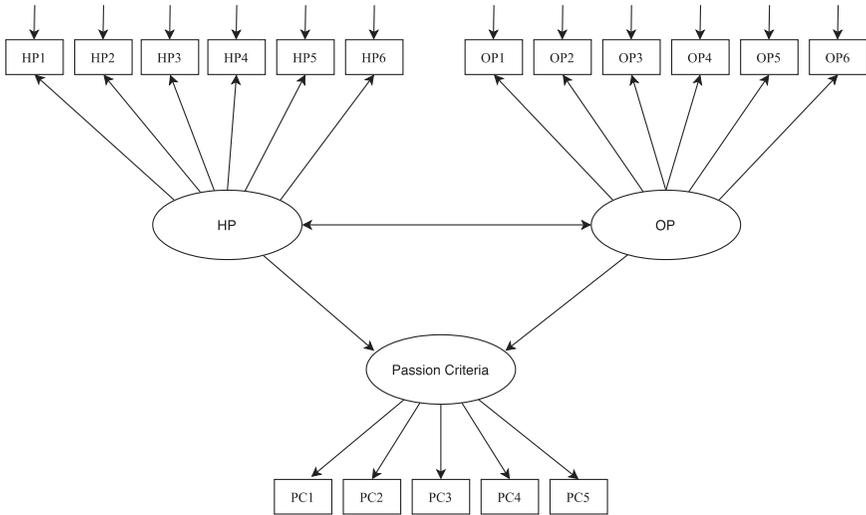


FIGURE 5. A structural model based on the CFA solution.
 Note: HP = harmonious passion; OP = obsessive passion; PC = passion criteria.

TABLE 6. CFA and ESEM structural models fit indices predicting L2 passion criteria

Model	χ^2	<i>p</i>	<i>df</i>	SRMR	RMSEA	RMSEA Low 90%	RMSEA High 90%	CFI	TLI
CFA	252.301	<.001	98	.09	.08	.07	.09	.89	.87
ESEM	187.299	<.001	89	.05	.06	.05	.08	.93	.91

TABLE 7. Standardized coefficients of HP and OP predicting L2 passion criteria through ESEM

Predicted variables passion criteria	ESEM		(uniq)	<i>R</i> ²
	HP	OP		
L2 Passion Criteria	.79**	.16*	(.31)	.69

*Path is significant at 0.05.

**Path is significant at the 0.01 level.

Note: (uniq) = item uniqueness; PC = L2 passion criteria; HP = harmonious passion; OP = obsessive passion.

DISCUSSION

The present study introduced the applicability and effectiveness of the relatively novel ESEM factor analysis method using the L2 Passion Scale. This investigation is probably the first of its kind in L2 research and has contributed to the enrichment of the literature on ESEM in general, while also opening the door for new directions in L2 psychometric examination research. The present study has provided evidence for the power of the

ESEM by comparing its results to those of its traditional CFA counterpart in two samples of language learners. Both studies' results show that ESEM resulted in a better fit to the data and generated a more realistic representation of the data based on responses to L2 Passion Scale among language learners using measurement invariance test. To extend the application of ESEM, the present study tested a structural ESEM-based model in which HP and OP were set as predictors of L2 passion criteria. The results illustrate the flexibility and usefulness of this novel method for SEM research.

CFA failed to provide an adequate fit to the data, even with the inclusion of the suggested correlation of uniqueness in the measurement model. This issue can be seen in past research in the field, leading researchers to apply several post hoc modifications in their measurement models, resulting in the distorted definition of their constructs (Collier, 2020; Hair et al., 2019; Marsh et al., 2009). This was evident in the study by Chen et al. (2021) with the Chinese version of the Passion Scale in which the researchers had to covary almost all measurement errors in the CFA model to improve the fit indices. Because of the overly restrictive requirement of the items to load on their perspective factors only, correlating the error terms was the most accessible approach to improve the model fit in the CFA. With multiple modifications included, one would argue that there is an unobserved factor that was failed to be accounted for in the measurement. Thus, providing empirical evidence, not theoretical support. Instead, ESEM can be, as the present study showed, a flexible method that overcome the need to apply unnecessary post hoc (*ex post facto*) modifications. Following previous DMP research (Horwood et al., 2021; Marsh et al., 2013; Tóth-Király et al., 2017; Vallerand, 2016), we believe that ESEM is a better fit to the DMP than the CFA as suggested by the goodness-of-fit and more realistic factor correlation. There was support for this *a priori* hypothesis for both language learners' samples.

In the ESEM solution, many cross-loadings were detected that, when not estimated, could erode the factorial structure of the L2 Passion Scale, as was the case in the CFA solution. Although two cross-loadings were significantly loaded on the other unrelated factor, they were not larger in size than the loadings on the intended factor. The first was OP3 "Learning English is the only thing that really turns me on," which loaded on its factor (i.e., OP), while loading on HP in Study 1 and Study 2, respectively. This can be justified through the wording of the items and the invariance of the solution across the two samples. It is possible that this item did not capture the specific aspect of OP exclusively; rather, it may have tapped into a more general, shared aspect of passion, resulting in significant loading onto the two factors. Moreover, perhaps the participants could understand this item's wording as, in part, a type of harmonious passion because their L2 learning could be perceived as harmonious but potentially as the only thing that attracts their attention when confronted with other activities. In contrast, OP1 "If I could, I would only do my English learning tasks" showed significant negative loading on HP, while showing significant positive loading on OP, thus indicating that this item might represent the two opposite poles of passion. For instance, a language learner with HP would openly engage in other important activities, demonstrating an ability to refrain from being involved in a controlled type of learning. The opposite is true for a language learner with OP. In summary, allowing the estimation of cross-loadings in ESEM provided a meaningful, flexible approach in L2 Passion Scale factor analysis (and possibly other scales), allowing for data interpretation from a novel perspective compared to the standard CFA (Horwood, 2021; Marsh et al., 2013; Tóth-Király et al., 2017).

We used ESEM to test the predictive power of HP and OP for L2 passion criteria. ESEM functioned in accordance with previous studies that have shown how the two types of passion relate to passion criteria. Of particular relevance, we found that HP outperformed OP in predicting the L2 passion criteria. These patterns of association are consistent with previous passion studies that examined the relationship between HP and OP, on the one hand, and passion criteria, on the other, demonstrating that HP is more strongly associated with these criteria than OP (Marsh et al., 2013; Vallerand, 2015; Zhao et al., 2015).

DIRECTIONS FOR L2 PSYCHOMETRIC RESEARCH

From a theoretical perspective, we encourage L2 researchers to replicate the findings of the L2 Passion Scale and L2 passion criteria and assess the extent to which they predict language learning outcomes. From a psychometric perspective, future research should replicate the factorial structure of the many L2 scales that have been solely established using EFA, standard CFA, or correlation by considering the application of ESEM and possibly contrasting competing measurement models. As the present research showed, ESEM can be extended to investigate the ability of the constructs to predict outcomes and contrast their predictive validity. This approach seems to be more rigorous and appropriate than using the widely implemented multiple regression analysis technique. This is because the ESEM structural model uses the rigor of SEM in both the measurement and prediction. In addition, researchers could consider the more recently developed forms of ESEM, namely, Bifactor ESEM (Alamer, 2021b, *in press*; Howard et al. 2018) and Set-ESEM (Marsh et al., 2020), and contrast the results to the standard ESEM. ESEM (and its variants) can be applied with Mplus or the “lavaan” R package. An applied example of ESEM can be found in the “Mplus User’s Guide,” which can be freely downloaded from the software website. To make this article more practical for analysts, we included the syntax required to run ESEM in the Mplus software (see [Appendix B](#)). We also make the raw data available for the researchers to replicate and rerun the analysis.

Further, the two studies discussed in this article have some limitations. First, although the EFA, CFA, and ESEM results yielded useful information with regard to the L2 Passion Scale’s construct validity, the correlated error terms of OP1 and OP2 in both Study 1 and Study 2 indicated that these two items share a common pattern of information that was not captured in the analyses. Further inspection into these two items could be useful. Second, although the number of participants in both studies satisfied the statistical guidelines for factor analysis (Hair et al., 2019; Kline, 2016), the sample sizes were relatively modest. Third, like other statistical analyses, ESEM, at its basic form infers group characteristics. Thus, researchers should consider multilevel modeling of ESEM as a compromise if hierarchical differences are believed to exist. In addition, ESEM is also subjected to rotational indeterminacy in that different rotation strategies result in different solutions that all fit the data adequately (Marsh et al., 2020). Fourth, although the ESEM appears to be less parsimonious than the corresponding CFA, some fit indices such as TLI and RMSEA control for parsimony (Marsh et al., 2009). However, we note that whilst TLI and RMSEA do control for parsimony, they are not the only control for parsimony. Ultimately, the balance between parsimony and goodness-of-fit should be based on evaluation of the parameter estimates in relation to theory, *a priori* predictions, evaluation of the parameter estimates. Lastly, it would be better for the assessment of discriminant validity to include

variables that may share similarities with HP and OP such as consistency of interest (from Grit) and intrinsic motivation (from Self-Determination Theory).

CONCLUSION

Overall, the present research introduced the ESEM method to the language learning and education field and applied it to the DMP framework. The study offered two scales to the field, the L2 Passion Scale and the L2 passion criteria and assessed them in a more advanced perspective. Using the advancement of ESEM, this study provided evidence that ESEM, as a synergistic product of EFA and CFA, is useful and powerful in testing the construct validity and unidimensionality of the L2 Passion Scale and possibly other L2 scales. In addition to construct validity, ESEM offered methodological features that allow researchers to evaluate convergent, discriminant, and predictive validity from a new empirical perspective (Guay et al., 2015), as this research has demonstrated. Moreover, the versatility of the ESEM, as it builds directly on SEM, can provide a more realistic, flexible representation of the already established instruments in the L2 domain. Thus, L2 researchers are encouraged to endorse ESEM and its extensions in their research to test and replicate the hypothetical factorial structures.

DATA AVAILABILITY STATEMENT

The experiment in this article earned Open Materials and Open Data badges for transparent practices. The materials and data are available at <https://www.iris-database.org/iris/app/home/detail?id=york:939777>.

REFERENCES

- Alamer, A. (2021a). Basic psychological needs, motivational orientations, effort, and vocabulary knowledge: A comprehensive model. *Studies in Second Language Acquisition*. Advance online publication. <https://doi.org/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*, 12, 732016. <https://doi.org/10.3389/fpsyg.2021.732016>
- Alamer, A. (2021c). Grit and language learning: Construct validation of grit and its relation to later vocabulary knowledge. *Educational Psychology*, 41, 544–562. <https://doi.org/10.1080/01443410.2020.1867076>
- Alamer, A., & Almulhim, F. (2021). The interrelation between language anxiety and self-determined motivation: A mixed methods approach. *Frontiers in Education*, 6, 618655. <https://doi.org/10.3389/educ.2021.618655>
- Alamer, A., & Lee, J. (2019). A motivational process model explaining L2 Saudi students' achievement of English. *System*, 87, 102133. <https://doi.org/10.1016/j.system.2019.102133>
- Alamer, A. (in press). Exploratory Structural Equation Modeling (ESEM) and Bifactor ESEM for construct validation purposes: Guidelines and example. *Research Methods in Applied Linguistics*.
- Ardasheva, Y., Tong, S., & Tretter, T. (2012). Validating the English language learner motivation scale (ELLMS): Pre-college to measure language learning motivational orientations among young ELLs. *Learning and Individual Differences*, 22, 473–483. <https://doi.org/10.1016/j.lindif.2012.03.001>
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 16, 397–438. <https://doi.org/10.1080/10705510903008204>
- Brown, K. W., & Ryan, R. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology*, 84, 822–848. <https://doi.org/10.1037/0022-3514.84.4.822>

- Chen, X., Vallerand, R. J., & Padilla, A. M. (2021). On the role of passion in second language learning and flourishing. *Journal of Happiness Studies*, 2761–2779. <https://doi.org/10.1007/s10902-020-00339-0>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233–255. https://doi.org/10.1207/S15328007SEM0902_5
- Collier, J. (2020). *Applied structural equation modeling using AMOS: Basic to advanced techniques*. Routledge. <https://doi.org/10.4324/9781003018414>
- Fernet, C., Lavigne, G. L., Vallerand, R. J., & Austin, S. (2014). Fired up with passion: Investigating how job autonomy and passion predict burnout at career start in teachers. *Work & Stress*, 28, 270–288. <https://doi.org/10.1080/02678373.2014.935524>
- Guay, F., Morin, A. J. S., Litalien, D., Valois, P., & Vallerand, R. J. (2015). Application of exploratory structural equation modeling to evaluate the academic motivation scale. *The Journal of Experimental Education*, 83, 51–82. <https://doi.org/10.1080/00220973.2013.876231>
- Hair, J., Black, W., Babin, B., & Anderson, R. (2019). *Multivariate data analysis*. 8th ed. Prentice Hall.
- Hancock, H., & Schoonen, R. (2015). Structural equation modeling: Possibilities for language learning researchers. *Language Learning*, 65, 160–184. <https://doi.org/10.1111/lang.12116>
- Horwood, M., Marsh, H. W., Parker, P. D., Riley, P., Guo, J., & Dicke, T. (2021). Burning passion, burning out: The passionate school principal, burnout, job satisfaction, and extending the dualistic model of passion. *Journal of Educational Psychology*, 113, 1668–1688. <http://doi.org/10.1037/edu0000664>
- Howard, L., Gagné, M., Morin, J., & Forest, J. (2018). Using bifactor exploratory structural equation modeling to test for a continuum structure of motivation. *Journal of Management*, 44, 2638–2664. <https://doi.org/10.1177/0149206316645653>
- Jamovi Project. (2019). *Jamovi*.
- Kline, R. (2016). *Principles and practice of structural equation modeling*. 6th ed. Guilford Publications.
- Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103, 391. <https://doi.org/10.1037/0033-2909.103.3.391>
- 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, 102–119. <https://doi.org/10.1080/00273171.2019.1602503>
- Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cut-off values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*, 11, 320–341. https://doi.org/10.1207/s15328007sem1103_2
- Marsh, H. W., Morin, A. J. S., Parker, P. D., & 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., 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, 439–476. <https://doi.org/10.1080/10705510903008220>
- Marsh, H. W., Vallerand, R. J., Lafrenière, M. K., Parker, P., Morin, A. J. S., Carbonneau, N., Jowett, S., Bureau, J. S., Fernet, C., Guay, F., Abduljabbar, A. S., & Paquet, Y. (2013). Passion: Does one scale fit all? Construct validity of two-factor Passion Scale and psychometric invariance over different activities and languages. *Psychological Assessment*, 25, 796–809. <https://doi.org/10.1037/a0032573>
- McDonald, R. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, 23, 1–21.
- Morin, A. J. S., Marsh, H. W., & Nagengast, B. (2013). Exploratory structural equation modeling. In *Structural equation modeling: A second course, 2nd ed., Quantitative methods in education and the behavioral Sciences: Issues, research, and teaching* (pp. 395–436). IAP Information Age Publishing.
- Morin, S., Arens, K., Tran, A., & Caci, H. (2016). Exploring sources of construct-relevant multidimensionality in psychiatric measurement: A tutorial and illustration using the composite scale of morningness. *International Journal of Methods in Psychiatric Research*, 25, 277–288. <https://doi.org/10.1002/mpr.1485>
- Muthén, K., & Muthén, O. (2014). *Mplus user's guide*. Muthén & Muthén.
- Ploonsky, L. 2015. *Advancing quantitative methods in second language research*. Routledge.
- Pollack, J. M., Ho, V. T., O'Boyle, E. H., & Kirkman, B. L. (2020). Passion at work: A meta-analysis of individual work outcomes. *Journal of Organizational Behavior*, 41, 311–331. <https://doi.org/10.1002/job.2434>

- Ryan, R., & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68–78. <https://doi.org/10.1037/110003-066X.55.1.68>
- Ryan, R., & Deci, E. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publications.
- Tóth-Király, I., Bőthe, B., Rigó, A., & Orosz, G. (2017). An illustration of the exploratory structural equation modeling (ESEM) framework on the Passion Scale. *Frontiers in Psychology*, 8, 1968. <https://doi.org/10.3389/fpsyg.2017.01968>
- Vallerand, R. J. (2015). *The psychology of passion: A dualistic model*. Oxford University Press.
- Vallerand, R. J. (2016). The dualistic model of passion: Theory, research, and implications for the field of education. In W. C. Liu, J. C. K. Wang, and R. M. Ryan (Eds.), *Building autonomous learners: Perspectives from research and practice using self-determination theory* (pp. 31–58). Springer Singapore.
- Vallerand, R. J., Blanchard, C., Mageau, G. A., Koestner, R., Ratelle, C., Léonard, M., Gagné, M., & Marsolais, J. (2003). Les passions de l'âme: On obsessive and harmonious passion. *Journal of Personality and Social Psychology*, 85, 756–767. <https://doi.org/10.1037/0022-3514.85.4.756>
- Verner-Filion, J., Vallerand, R. J., Amiot, C. E., & Mocanu, I. (2017). The two roads from passion to sport performance and psychological well-being: The mediating role of need satisfaction, deliberate practice, and achievement goals. *Psychology of Sport and Exercise*, 30, 19–29. <https://doi.org/10.1016/j.psychsport.2017.01.009>
- Woodrow, L. (2014). *Writing about quantitative research in applied linguistics*. Palgrave Macmillan.
- Zhao, Y., St-Louis, A., & Vallerand, R. J. (2015). On the validation of the Passion Scale in Chinese. *Psychology of Well-Being*, 5, 3. <https://doi.org/10.1186/s13612-015-0031-1>

APPENDIX A

TABLE A1. L2 Passion Scale items and L2 passion criteria

Item	Factor
Learning English is in harmony with the other activities in my life.	HP1
The new things that I discover in English allow me to appreciate it even more.	HP2
Learning English reflects the qualities I like about myself.	HP3
Learning English allows me to live a variety of experiences.	HP4
Learning English is well-integrated in my life.	HP5
Learning English is in harmony with other things that are part of me.	HP6
If I could, I would only do my English learning tasks.	OP1
I have almost an obsessive feeling for learning English.	OP2
Learning English is the only thing that really turns me on.	OP3
I have difficulties controlling my urge to learn English.	OP4
Learning English is so exciting that I sometimes lose control over it.	OP5
I have the impression that learning English controls me.	OP6
<i>L2 Passion Criteria</i>	
I spend a lot of time engaging in my English learning.	PC1
I like my English learning.	PC2
Learning English is important to me.	PC3
Learning English is my passion.	PC4
Learning English is part of who I am.	PC5

Note: HP = harmonious passion; OP = obsessive passion; PC = L2 passion criteria.

APPENDIX B

```
=====
VARIABLE:
  NAMES = Gender   HP1   HP2   HP3   HP4   HP5   HP6
         OP1   OP2   OP3   OP4   OP5   OP6;
  MISSING = ALL (-99);

USEVARIABLES = HP1   HP2   HP3   HP4   HP5   HP6
              OP1   OP2   OP3   OP4   OP5   OP6;

MODEL: HP OP by HP1-OP6(*1);

!For error covariance:
  OP2 WITH OP1;

ESTIMATOR IS MLR;
OUTPUT: STDYX;
```

FIGURE B1. Mplus syntax (code) for running ESEM with the data of Study 1.

```
=====
VARIABLE:
  NAMES = Gender   HP1   HP2   HP3   HP4   HP5   HP6
         OP1   OP2   OP3   OP4   OP5   OP6
         PC1   PC2   PC3   PC4   PC5;

USEVARIABLES = HP1   HP2   HP3   HP4   HP5   HP6
              OP1   OP2   OP3   OP4   OP5   OP6

MODEL:
  HP OP by HP1-OP6(*1);
  PC by PC1-PC5;
  PC on HP OP;

!For error covariance:
  OP2 WITH OP1;

ESTIMATOR IS MLR;
OUTPUT: STDYX;
```

FIGURE B2. Mplus syntax (code) for running structural ESEM.

```

=====
VARIABLE:
    NAMES = Gender    HP1  HP2  HP3  HP4  HP5  HP6
           OP1  OP2  OP3  OP4  OP5  OP6;

USEVARIABLES =  HP1    HP2  HP3  HP4  HP5  HP6
                OP1  OP2  OP3  OP4  OP5  OP6

MODEL:
    f1 by HP1-HP6;
    f2 by OP1-OP6;

!For error covariance:
OP1 WITH OP2;

ESTIMATOR IS MLR;
OUTPUT: STDYX;

```

FIGURE B3. Mplus syntax (code) for running CFA.

APPENDIX C: EXPLORATORY FACTOR ANALYSIS

An EFA with Oblimin rotation was selected because the DMP factors tap similar aspects of passion. EFA was performed using Jamovi software 1.1 (Jamovi Project, 2019) to identify information for induction into the L2 Passion Scale’s possible factor structure. We posited *a priori* that there would be two well-defined factors. We tested this by evaluating (a) eigenvalues greater than 1, (b) the scree plot of the eigenvalues, (c) factor loadings greater than .30, and (d) the substantive interpretability of the factor structures. As shown in Table 1 and Figure 3, the results yielded a two-factor structure for HP and OP, accounting for 59.60% of the variance. These two factors reflected the L2 Passion Scale’s hypothesized factorial structure.

TABLE B1. EFA factor loadings of HP and OP based on Oblimin rotation

	Factor		Uniqueness
	1	2	
HP1		.44	.80
HP2		.66	.56
HP3		.65	.55
HP4		.71	.49
HP5		.65	.57
HP6		.48	.76
OP1	.65		.55
OP2	.68		.54
OP3	.44	.38	.64
OP4	.47		.77
OP5	.52		.61
OP6	.61		.56

Note: The minimum residual extraction method was used in combination with an Oblimin rotation.

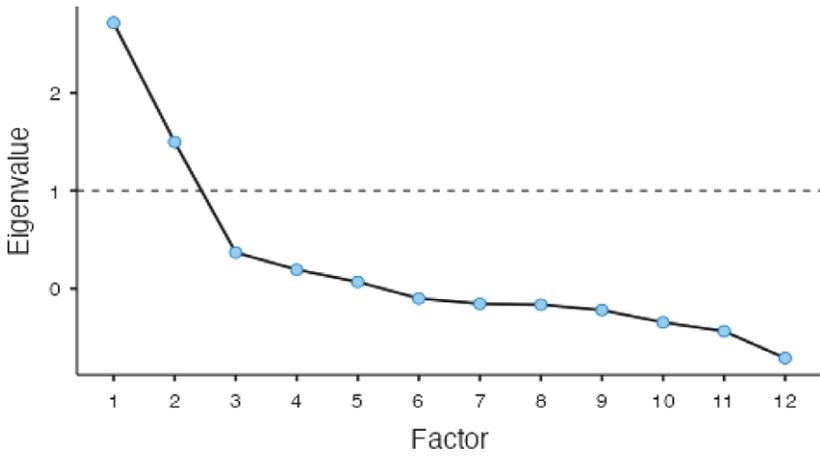


FIGURE B4. A scree plot based on the Oblimin solution.