Research Methods in Applied Linguistics 1 (2022) 100027

Contents lists available at ScienceDirect

Research Methods in Applied Linguistics

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

Partial Least Squares Structural Equation Modeling (PLS-SEM) in second language and education research: Guidelines using an applied example

Joseph Hair[a], Abdullah Alamer[b,c,*]

[a] Mitchell College of Business, University of South Alabama, USA
[b] King Faisal University, Saudi Arabia
[c] The University of New South Wales (UNSW), School of Education, Australia

A R T I C L E I N F O

Keywords:
Partial least squares structural equation modeling (PLS-SEM)
Covariance-based structural equation modeling (CB-SEM)
Structural equation modeling (SEM)
Convergent validity
Discriminant validity
Predictive power

A B S T R A C T

Partial least squares structural equation modeling (PLS-SEM) is an alternative method to the historically more commonly used covariance-based SEM (CB-SEM) when analyzing the data using structural equation modeling (SEM). The article starts by introducing PLS-SEM to second language and education research, followed by a discussion of situations in which PLS-SEM should be the method of choice for structural equation modeling. It is argued that PLS-SEM is appropriate when complex models are analyzed, when prediction is the focus of the research – particularly out-of-sample prediction to support external validity, when data do not meet normal distribution assumptions, when formative constructs are included, and when higher-order constructs facilitate better understanding of theoretical models. The most up-to-date guidelines for applying PLS-SEM are provided, and step-by-step guidance is offered on how to apply the method using an R statistical package (i.e., SEMinR) that is available. An example is provided that shows how the results of PLS-SEM are interpreted and reported. We also make the data publicly available for readers to start learning PLS-SEM by replicating our findings. The paper concludes with important considerations for the utilization of SEM, especially PLS-SEM, in future L2 research.

Introduction

Structural equation modeling (SEM) is a very useful technique for evaluating complex theoretical relationships between multiple variables, especially when conducting social science and second language (L2) research. Two fundamental SEM methods have been proposed: covariance-based structural equation modeling (CB-SEM) and partial least squares structural equation modeling (PLS-SEM, also referred to as composite-based structural equation modeling). CB-SEM has been well-known in the L2 field for many years and is by far the most dominant statistical technique among researchers conducting SEM (Shao et al., 2022).

PLS-SEM is being applied more often recently as more researchers have become aware of the method (Hair et al., 2019a; Hair et al., 2022a; Ringle et al., 2015; Sarstedt et al., 2020). PLS-SEM is a suitable alternative to CB-SEM in a number of situations that are common in L2 quantitative research, and thus may be a more appropriate method to use, as will be elaborated on later. The number

Abbreviations: PLS-SEM, Partial least squares structural equation modeling; CB-SEM, Covariance-based structural equation modelling; ESEM, Exploratory structural equation modeling; CFA, Confirmatory factor analysis; SEM, Strcutural equation modeling; CI, Consistency of nteresti; PE, Perseverance of effort.

* Corresponding author.
E-mail address: alamer.aaa@gmail.com (A. Alamer).

https://doi.org/10.1016/j.rmal.2022.100027
Received 4 May 2022; Received in revised form 15 July 2022; Accepted 16 July 2022

of empirical studies using PLS-SEM in the language learning domain is still small, however, and to date, very few studies have opted to use this approach for executing SEM (Alamer, 2022b; Elahi Shrivan & Alamer, 2022; Law & Fong, 2020; Mohammadi et al., 2020). Although a plethora of publications in research fields other than language learning have addressed the use of PLS-SEM (particularly in marketing and information systems), there is no introductory paper explaining the features and key aspects of this method in the L2 and education research. Moreover, there is no illustrative example for the stages involved in assessing PLS-SEM results according to the most recent developments in the field.

The question remains: Why should L2 and education researchers consider using the alternative structural equation modeling approach of PLS-SEM when an established approach of CB-SEM has been applied for the previous 40+ years to assess theoretical models? Answering this question is the primary objective of this article. The article begins by describing the two methods of structural equation modeling, explaining how the two approaches differ, and the procedure scholars can follow in deciding which of the methods is appropriate to achieve their research objective. To facilitate scholarly decision making, we provide established guidelines that clearly delineate how and when to use each method. To make this article more practical for application in the field, a step-by-step guide for evaluating PLS-SEM results is presented, using the free, open-source statistical package SEMinR. We explain the steps while assuming the reader already has basic knowledge of SEM and is seeking to understand the specific details of the more recently emerging PLS-SEM method. The empirical example we use is drawn from the L2 motivation literature (Alamer, 2021b, Alamer, 2022d; Elahi Shrivan & Alamer, 2022); therefore, it should be relevant for L2 researchers in general.

Structural equation modeling: the two methods

There are two broad approaches to executing SEM. One approach is covariance-based SEM, popularized in the early 1980's by Jöreskog (1970) and Jöreskog and Sorbom (1976). The other broad SEM approach is variance-based SEM (also called composite-based SEM). Among variance-based SEM, partial least squares structural equation modeling (PLS-SEM) is regarded as a fully developed and general approach (Hair et al., 2019a). The PLS-SEM approach most often applied in social sciences research is the one originally proposed by Wold (1982) and later popularized by Chin (1998) and more recently by Hair et al. (2011) and Ringle et al. (2015). Indeed, several researchers have noted that the PLS-SEM has attracted considerable attention in social sciences research in the last decade (Chin et al., 2020; Hair et al., 2022a; Hair et al., 2019a; Law & Fong, 2020; Sarstedt et al., 2014).

The CB-SEM and PLS-SEM approaches were initially proposed around the same time (i.e., the 1960–70s), but virtually all social sciences applications of SEM were largely conducted through covariance-based SEM until the past decade. Numerous CB-SEM statistical software tools, such as AMOS, EQS, LISREL, M*plus,* and SAS, among others, emerged in the 1980's and 1990's, and the method dominated the SEM field until around 2012. The software tools such as PLSgraph (Chin, 1994) and SmartPLS (Ringle et al., 2005, 2015), as well as the SEMinR free software package for R in particular, have enabled many more researchers to expand their methodological toolbox to pursue alternative research objectives that were restricted by the limitations of the CB-SEM approach.

CB-SEM is primarily designed to test established theoretical frameworks, including both the measurement and structural models (e.g., examining whether theorized relationships are present in the data). This is achieved by comparing the estimated theoretical model (covariance matrix) with the observed covariance matrix (Hair et al., 2019a; Hair et al., 2022b). The focus, therefore, is on the explanation of the relationships between all the variables in the analysis (i.e., theory confirmation). An example of such approach is the flexibility of CB-SEM in theory testing by the applications of exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and the most recently developed technique of exploratory structural equation modeling (ESEM) in the examination of the multidimensionality of a theoretical construct (Alamer, 2022c, 2021a; Alamer & Marsh, 2022; Shao et al., 2022).

Another characteristic of CB-SEM is that the method is referred to as a factor-based approach that assumes a common factor model. That is, the covariances between the indicators are only the product of (caused by) the factor. The PLS-SEM method, in contrast, is a composite-based approach that uses total variance (common, specific, and error variance) and represents the construct as a linear combination of its indicators (Hair et al., 2022a; Hair & Sarstedt, 2021; Lohmöller, 1989; Sarstedt et al., 2016).

Finally, CB-SEM solutions require more rigorous assumptions, including multivariate normality, and larger sample sizes with the maximum likelihood (ML) estimator. Although some estimators such as the robust maximum likelihood (referred to as MLM and MLR) can be used to account for the violation of multivariate normality in the CB-SEM (Kline, 2016; Shao et al., 2022), MLR is available as an option for only a few SEM packages; for example, it is not incorporated into Amos and EQS versions older than 6. Although estimators other than ML that take the nonnormality of the data into account are available (such as weighted least squares and diagonally weighted least squares), SEM literature largely favors ML over any other estimators because of its asymptotic unbiasedness and consistency in estimating the model fit indices (Kline, 2016). In contrast, the PLS estimator does not assume the normality of the data by default. Note that some researchers advocate for transforming the data when it does not satisfy the normality assumption in SEM applications. We do not recommend this procedure, however, as it may alter the interpretation of the variable (Hair et al., 2019). When SEM estimators that are robust to nonnormality are available, they should be prioritized.

PLS-SEM is especially useful when the user's structural model objective is to predict and explain the target outcomes as obtained by the in-sample and out-of-sample metrics (Latan & Noonan, 2017; Shmueli et al., 2019; Hair et al., 2019a; Hair et al., 2022a; Hair & Sarstedt, 2021). PLS-SEM also obtains solutions with smaller sample sizes compared to CB-SEM. How is that possible, particularly compared to CB-SEM? PLS-SEM iterates back and forth multiple times optimizing first the measurement model and then the structural model, and back to the measurement model, and then again to the structural model, and so on, continuing until the ultimate objective of optimizing prediction, and not model fit, is achieved. In fact, the name "partial least squares" was derived from the "partial" approach to analyzing data, as compared to CB-SEM, which analyzes all the data simultaneously to obtain solutions and therefore requires considerably larger sample sizes. The ability to obtain solutions with smaller sample sizes is a result of this partial approach to analyzing the data. When possible, however, larger sample sizes should always be used to substantiate the ability to infer sample results to the relevant population (Alamer, 2022c; Hair et al., 2022b; Kock & Hadaya, 2018). Finally, some scholars have argued that CB-SEM is better than composite-based SEM methods since composite-based SEM methods supposedly do not account for measurement error (Ronkko et al., 2015). Nevertheless, Hair and Sarstedt (2021) have demonstrated that PLS-SEM does in fact account for measurement error.

Characteristics of PLS-SEM

Applications of PLS-SEM have increased substantially in the social sciences in general, and several studies have made use of the potential of PLS-SEM in L2 and education research (e.g., Alamer, 2022b, 2022d; Elahi Shrivan & Alamer, 2022; Law & Fong, 2020). When deciding whether PLS-SEM is the appropriate method for the research, one should consider the following aspects (Hair et al., 2022a):

- Objective. To maximize the explained variance in the outcome variable(s), including assessment of both in-sample and out-of-sample prediction. Thus, it is preferable when explanation plus prediction are the main objectives in the structural model (Shmueli et al., 2019).
- Measurement model types. PLS-SEM easily handles formative constructs (referred to as composite)– constructs with arrows pointing from the observable variables without posing specific constraints on the model. In fact, we believe this is a key feature that might motivate L2 researchers to select PLS-SEM over CB-SEM. We elaborate on formative constructs in a later section.
- Sample size. Although larger sample sizes are always preferred, compared to CB-SEM, PLS-SEM obtains results with smaller sample sizes and achieves high levels of statistical power (Hair et al., 2017). Notwithstanding, researchers should consult guidelines about sample size recommendations such as Kock and Hadaya (2018) and Hair et al. (2022a).
- Distribution. PLS-SEM does not assume normally distributed data and is quite robust to skewness.
- Model complexity. PLS-SEM is robust to quite complex models containing literally hundreds of observed variables and rarely faces convergence issues.
- The mainstream PLS-SEM algorithm. It does not specify goodness-of-fit (model fit indices) as an assessment metric when obtaining structural model solutions. Instead, the investigator relies on a different set of indices, including construct reliability and validity, as well as in-sample and out-of-sample prediction indices.
- Higher (e.g., second) order constructs (HOCs). The method can easily obtain solutions with as few as two lower (first) order constructs (LOCs). It also allows for LOCs to be formulated as formative for the HOC, which is not possible in CB-SEM without a compromise in model specification (Sarstedt et al., 2019).

As noted above, there are several reasons to consider when determining whether to use PLS-SEM. Probably the most appealing characteristics of PLS-SEM in the L2 and education domain are: (i) the method does not require normally distributed data by default; (ii) it obtains solutions with smaller sample sizes compared to CB-SEM; and (iii) it easily incorporates formatively (composite) measured constructs. Specifically, (i) it is important to note that using non-normally distributed data with CB-SEM may result in inflated, thus biased, model parameters when the researcher incorrectly assumes the normality of non-normal data through the default ML estimation (see the discussion in an earlier section). In addition, (ii) in many situations, the focus of the L2 research is placed on understanding the characteristics of language teachers (or parents for education research), which we know from our experience, are limited in number as part of the sample compared to students. In such cases, PLS-SEM can be a suitable alternative that overcomes typical sample size limitations. Specifically, it is estimated that to obtain significant results ($p < .05$) for path coefficients ranging between .21 and .30, the user may be able to achieve this with a sample size aound 69 to 100, though larger sample size is typically favored (Hair et al., 2022a; Kock & Hadaya, 2018). Furthermore, (iii) language researchers often deal with formative constructs (composite) in which the items form the construct in linear combinations are not assumed to be highly correlated. For example, 'language achievement' as a construct usually consists of the four skills: listening, speaking, reading, and writing (among others). In such cases, the construct should *not* be specified as reflective (which is almost always specified this way in the field), but rather as formative. This is because the four language skills are not interchangeable and can be different at the learner level. For example, one learner could score high in writing but low in speaking. So, high correlations between the four skills are not feasible. In contrast, reflective constructs are those typically employed in CB-SEM such as in the CFA and ESEM applications in which the construct explains a set of observed variables and not the other way around. As such, a change in one item should not influence the overall meaning of the construct.

The same observation can be made with higher-order factors in which the first-order factors may be better conceptualized as formative than reflective. For example, consider three specific language reading strategies that are hypothesized to belong to a global (higher order) *reading strategy factor*. The three types of reading strategies are theorized to function independently from each other, such that high correlations between them may not necessarily be observed (see, for example, the metacognitive reading strategies identified in Mokhtari and Reichard (2002)). In such cases, the three specific reading strategies should be specified formatively relative to their higher-order (global) factor. Hence, when constructs are more appropriately structured as higher order measurement models, as is true with the language achievement construct and reading strategies, an advantage of PLS-SEM is that postulating this type of construct is quite easily set up with PLS (Sarstedt et al., 2019). In contrast, examining this type of construct is much more challenging with CB-SEM since researchers will be confronted by identification issues with this type of formulation. Moreover, this difficulty can only be resolved by inserting additional model constraints to the global factor in the model (also called MIMIC model), which sometimes compromises the substantive theory. The ease of configuration of formatively measured constructs in PLS-SEM is a

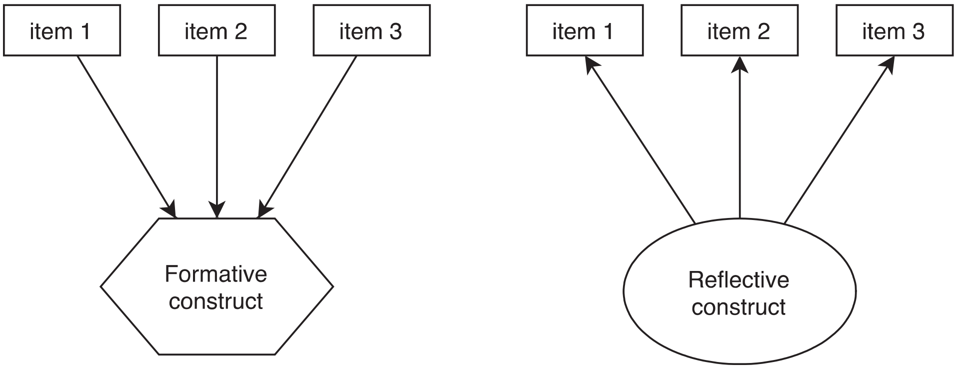

Fig. 1. A visual example of formative and reflective constructs.
Note. Hexagon shape denotes a formative construct

definite advantage since this type of construct is commonly encountered in L2 research but often inappropriately operationalized in empirical studies. See Fig. 1 for conceptual differences between the two types of constructs.

Previous studies applied PLS-SEM in the L2 domain

In the language learning domain, research has started to make use of these key features of PLS-SEM. For instance, Sparks and Alamer (2022) have investigated the impact of L1 language skills on L2 achievement through different mediators. Achievement was measured through the four areas of reading, writing, speaking, and listening comprehension. The operationalization of *achievement* as a formative construct was not possible to be estimated through the CB-SEM in their structural model due to identification issues. Hence, selecting PLS-SEM was a suitable choice when formative constructs are involved. Another example can be taken from Alamer's (2022d) study as the research objective was to examine the predictive power of the *autonomous single language interest* (ASLI) of students' L2 achievement after a whole academic year. ASLI postulates that L2 students who have one autonomous language interest during their study would be able to achieve learning the language more than students with lower ASLI. To fully examine the predictive capability of ASLI, a structural model was hypothesized by applying the PLS-SEM method to achieve prediction purpose. Through the analysis of the in-sample and out-of-sample assessments (explained in a later section) the author found that ASLI successfully predicted unseen scores on L2 achievement indicating good prediction power which supported the external validity of the research findings from a new perspective (Alamer, 2022d). Because the investigation followed a causal–predictive paradigm, PLS-SEM was seen as more appropriate than CB-SEM for such an objective.

An empirical example

In our example, 213 Saudi participants who learn English as an L2 in a public Saudi university participated in an online questionnaire. All students, male and female, speak Arabic as their mother tongue, and their ages were between 18 to 26 (M_{age} = 20.3). The measures of the study are based on a 5-point Likert-type response format ranging from 1 (strongly disagree) to 5 (strongly agree), and full details can be obtained from Appendix B.

Brief overview of the constructs

To contextualize our example, Fig. 2 presents a hypothetical theoretical model based on the L2 motivation literature. The literature indicates two constructs, consistency of interest (CI) and perseverance of effort (PE), are the main components of the grit theory (Duckworth et al., 2007) and are theoretically positively associated with language outcomes (Elahi Shirvan & Alamer, 2022). At the same time, research has illustrated the limited predictive power of the two constructs for language related outcomes which may be due to the lack of understanding of the CI and PE antecedents (Alamer, 2022d). For example, recognizing the ways in which CI and PE can be sustained and enhanced among L2 students can be a fruitful endeavor to obtain better understanding of the psychological processes (Alamer, 2021b). A proposed possibility to fill this knowledge gap is the assessment of L2 students' basic psychological needs of autonomy, competence, and relatedness (Ryan & Deci, 2017) in which the literature highlights their role as a basis for sustaining autonomous motivation, adaptive functioning, and improved learning effort (Alamer, 2022a, 2022c; Alamer & Lee, 2019), even during difficult times such as the COVID-19 pandemic (Alamer, 2022b, Alamer & Al Sultan, 2022). It is important, therefore, to evaluate the extent to which the BPN operate as antecedents of CI and PE, which, in turn, lead to increased language achievement.

First, the basic psychological needs of autonomy, competence, and relatedness, are postulated as exogenous latent variables. Next, grit, which consists of CI and PE are postulated as endogenous constructs and mediators for the basic psychological needs (readers are referred to Alamer (2021b), and Elahi Shirvan and Alamer (2022) for greater details). Lastly, students' achievement is postulated as an endogenous composite variable and is also the outcome in the model. See Appendix B for the scale items for the basic psychological needs as well as L2 grit. Fig. 2 illustrates additional details on the specific relationships. The order of the constructs should be determined based on the available literature. Specifically, research in language learning shows that BPN are seen as the basis of a

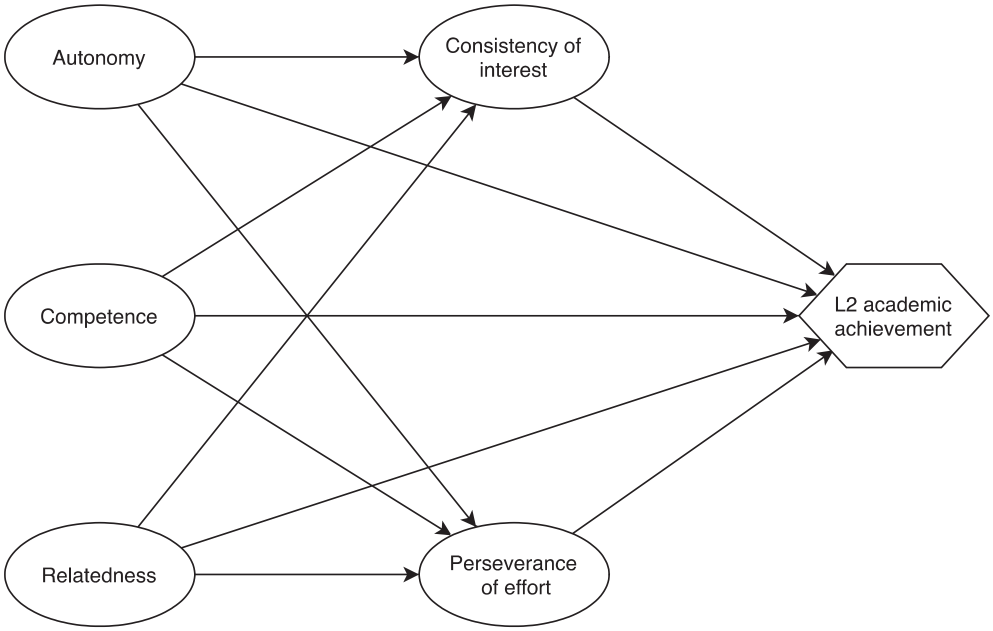

Fig. 2. The hypothesized relationships in the structural model.

learner's motivational orientation (Alamer, 2022a; Alamer & Almulhim, 2021; Noels et al., 2000; Noels, 2013; Oga-Baldwin et al., 2017).

Note: constructs represented in hexagon denote formative constructs (composite).

Steps for conducting PLS-SEM through SEMinR using the empirical example

Structural equation modeling, particularly PLS-SEM, is an increasingly popular analytical method. The R programming language is the language of choice for many researchers. R is a free, open-source software that enables users to write and execute code to analyze data. We use the SEMinR package for PLS-SEM (Ray et al., 2020; Hair et al., 2022b) to execute the empirical example in this article. Note that in the PLS literature measurement models are often referred to as the "outer" model while the structural model is identified as the "inner" model (Hair et al., 2022a).

Step 1: Downloading and installing the SEMinR package

To run PLS-SEM in the R environment, one can use the free SEMinR package. SEMinR requires at least the 2020 version of R (R Core Team, 2020). Novice readers can learn more about using R in L2 research by perusing primers such as Larson-Hall (2015) and Winter (2019). The SEMinR package is on the Comprehensive R Archive Network (CRAN) and is therefore available for direct download in the R software using the following code:

```
install.packages(seminr)
```

SEMinR executes all major features of PLS-SEM. SEMinR can be loaded using the following code:

```
library(seminr)
```

Step 2: Specifying the structural (inner) model

Typically, SEM begins by specifying the overall theoretical model, which includes two major parts: the structural model and the measurement model. Similar to CB-SEM, researchers using PLS-SEM usually start the modeling process by identifying their constructs and then postulate the linkages between the constructs. To do so, researchers must specify the sequence of the constructs in the model. In any structural model, there are exogenous variables (affecting another construct) and endogenous constructs (affected by another variable). Mediators are also endogenous constructs that have effects on other endogenous constructs in the model.

The structural model shown in Fig. 2 can be specified using the following code:

```
sm <- relationships(
paths(from = "AUT", to = c("PE", "CI", "Achievement")),
paths(from = "COM", to = c("PE", "CI", "Achievement")),
paths(from = "REL", to = c("PE", "CI", "Achievement")),
paths(from = "CI", to = ("Achievement")),
```

```
paths(from = "PE", to = "Achievement")
```

Note. AUT = autonomy, COM = competence, REL = relatedness, CI = consistency of interest, PE = perseverance of effort

Step 3: Specifying the measurement (outer) model

For the measurement model, researchers decide the nature of the relationship between the constructs and their indicators. Constructs can be reflective (i.e., the construct points towards its items) or formative (i.e., the items point towards their construct), depending on the theoretical measurement formulation of the construct. For further details about which type best represents the construct, the reader is referred to Hair et al. (2017; 2022a). As a quick tip, if the construct's items are postulated to reflect one specific concept, and they share similarities in meaning, then the construct may be better formulated as reflective. In contrast, if the items are conceptualized as a set of components that form a whole, and each item stands alone (i.e., they are not interchangeable in meaning), then the construct may be better formulated as formative (also referred to as composite measurement). For instance, if the construct is conceived to capture L2 proficiency and is assessed using learners' scores for different language skills, such as comprehension, speaking, vocabulary knowledge, and reading, then the construct should be operationalized as formative (or composite). Note that each construct has its own indicators. For example, autonomy, competence, and relatedness each has four items, taken from BPN-L2 scale (Alamer, 2022a). Example items are as follows: for autonomy, 'I can 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'. Similarly, CI and PE each has six items, taken from Alamer (2021b). Examples of items are as follows: for CI, 'I often set a language learning goal but later choose to pursue a different one' (reversed); and for PE, 'I work hard towards my language learning goals irrespective of how long they take to achieve'. All these constructs are modeled as reflective, but the outcome 'Academic achievement' is modeled as formative and consists of two items: (i) students' grade point average [GPA] in their L2 English program at university, and (ii) teacher's evaluation of students' language levels which was out of five points. There are 213 cases included in the analysis. We now specify the measurement models in the SEMinR environment using the following code:

```
mm <- constructs(
composite("CI", multi_items("CI", 1:6)),
composite("PE", multi_items("PE", 1:6)),
composite("AUT", multi_items("aut", 1:4)),
composite("COM", multi_items("com", 1:4)),
composite("REL", multi_items("rel", 1:4)),
composite("Achievement", multi_items("ach", 1:2), weights = mode_B))
```

Note. 'Achievement' is formulated as a formative construct which is indicated by 'mode B'.

The researcher can then load their data and prepare to run the model. We are now ready to run the PLS-SEM and assess its results.

Step 4: Assessing the PLS-SEM

Part 1: Assessing the (outer) model

Assessing reflective constructs is achieved by considering five steps: (1) estimate the loadings and their p-value, (2) estimate indicator reliability, (3) examine internal consistency reliability, (4) obtain the average variance extracted (AVE), (5) check the discriminant validity through HTMT (Hair et al., 2020).

When theoretical constructs are measured formatively, however, an alternative approach should be used to test their validity: (1) estimate the convergent validity, (2) check indicator multicollinearity, (3) test the size and significance of indicator weights, (4) examine the size and significance of indicator loadings, and (5) assess predictive validity (Hair et al., 2022). In assessing reflective constructs, the researcher should consider the following steps:

(1) Estimate the loadings and their p-values. It is suggested that indicator loadings should be in the region of .70 and statistically significant at .05 or below (equal to t-statistic of $\pm$ 1.96). However, values between .40 and .70 can be justified (Hair et al., 2019) if acceptable values are obtained on other indices (e.g., indices obtained from steps 3, 4, and 5). Note that p-values of the loadings in PLS-SEM are typically obtained by running a bootstrapping procedure (Hair et al., 2013). Researchers can also check the 95% confidence intervals (CI) to determine the range 95% of the estimated loadings will fall within when repeatedly sampling from the population of interest randomly. Research indicates that bias-corrected and accelerated (BCa) bootstrap approach adjusts for biases and skewness in the data, and thus should be used when the data depart from normality (Hair et al., 2017). Moreover, confidence intervals that do not include a zero indicate statistical significance. Values below the cut-off value should not be removed automatically. Rather, the researcher needs to check that such removal does not affect convergent validity and reliability.
(2) Estimate indicator reliability. Indictor reliability can be obtained by squaring the individual indicator loadings. A value of .50 is considered acceptable to indicate the item shares at least 50% of variance with the construct.
(3) Examine construct internal consistency reliability. Internal consistency reliability is assessed using Cronbach's alpha (α) and Composite Reliability (CR). The cut-off value of .70 for both measures appear to be acceptable and is widely applied in PLS-SEM research (Hair et al., 2017). Given that the CR value is often less restrictive than α, it can be used as the upper band, with the α value as the lower band (Hair et al., 2018). Nonetheless, if the reliability estimate is .95 or higher, this indicates individual items are possibly measuring essentially the same aspect of the construct, and thus they are redundant (Hair et al., 2022a).

(4) **Obtain the Average Variance Extracted (AVE)**. This measure refers to the extent to which items on a specific construct correlate positively and share a high degree of variance as assessed through AVE. The rule of thumb is that values of .50 or higher provide evidence of the convergent validity of the construct. Mathematically, a value of .50 indicates the mean values of the items' factor loadings are .708 or higher, thus indicating a sufficiently meaningful relationship between the variances of the items and their presumed construct.
(5) **Check the discriminant validity through HTMT.** This measure reflects the extent to which the construct is conceptually different from other constructs involved in the study. In other words, each of the theoretical constructs in the model is unidimensional (measures a uniquely different concept and exhibits only a small amount of overlap in variances). The recommended measure of discriminant validity is heterotrait-monotrait ratio (HTMT) of the correlations. The HTMT ratio reflects the extent to which a construct better explains the variance in its own indicators, compared to the variance of other constructs. A conservative cut-off value of HTMT < .85 can be used, but if there are possible communalities among the indicators on the constructs, then the researcher can use the more liberal cut-off value < .90 (Hair et al., 2022a).

When the conceptual model includes *formative constructs* (such as 'achievement' in our case), the procedure suggests different steps to assess these types of constructs in the model (Hair et al., 2020; Hair & Sarstedt, 2020). Because the observed variables in a formative construct are not theorized to be correlated, the assessment does not include steps that evaluate the strength of the interrelationships between the constructs (communality). In assessing formative constructs, the researcher should consider the following steps.

(1) **Evaluate the convergent validity.** This step differs from convergent validity assessment in reflective constructs. The assessment evaluates the extent to which the substantive construct is positively associated with a reflective measure of the same concept. Note that the substantive construct could be negatively associated with the specified measure if theory indicates such a relationship. In addition, the benchmark construct can be either a single item or multiple items depending on the complexity of the construct. Thus, the benchmark measure should be considered early in the data collection. For example, an L2 researcher would expect that achievement would be correlated with a reflective (or single item) measure of language engagement.
(2) **Examine indicator multicollinearity.** In contrast to reflective constructs, formative constructs assume items within a construct are not highly correlated, and therefore they are not interchangeable. The suggested measure of collinearity is the variance inflation factor (VIF), which is the reciprocal of tolerance. It is suggested that a VIF value of 5 or greater indicates serious collinearity issues among the predictor constructs (Hair et al., 2018; 2021). VIF values below 3 imply no collinearity, but values between 5 and 3 can be acceptable, with theoretical justifications. When high multicollinearity is present, the researcher can check if one of the items can be removed. Nevertheless, because each item in formative constructs holds unique information, this action should only be taken with caution. Possibly, the multicollinearity may suggest the need for the inclusion of a second-order construct (HOC) that is conceptually supported by substantive theory. Alternatively, the researcher can use Mode A for the formative construct if collinearity is present and theoretically justified.
(3) **Assess the size and significance of the indicator weights.** In this step, we evaluate the robustness of the observed variables in contributing to and forming the construct. The *weights* of the items in formative constructs are similar to the beta coefficients in multiple regression model. Therefore, the weights are usually smaller in size than when relationships are estimated as reflective. Note that weights are typically assessed as statistically significant at $p \leq .05$, but $\leq .10$ is acceptable with smaller sample sizes. In addition, the size of the weights should be considered, with larger indicator weights representing an individual indicator contributing more (relevance) to the construct.
(4) **Evaluate the indicators loadings (additional).** In cases where an item on a formative construct is not significant, the researcher should not automatically remove the item. Instead, the item should be evaluated based on the absolute contribution of the item by inspecting its *loading.* If the item loading is statistically significant and ≥ .50 in magnitude, this empirically justifies retaining the item because it makes a sufficient absolute contribution to forming the construct (Hair et al., 2022b).

Part 2: Assessment of the structural (inner) model

Having established the reliability and validity of the constructs, we now examine the structural component of the model: the structural or inner model. Since the PLS-SEM algorithm is not based on the variance-covariance matrix, researchers should assess the structural model based on its ability to predict the outcome(s). Recent guidelines (Hair et al., 2020) recommend the following steps to assess the structural model: (1) examine the model for collinearity; (2) evaluate the size and significance of the paths; (3) assess the coefficient of determination (R^2); (4) examine out-of-sample predictive power, using the $\text{PLS}_{\text{predict}}$ method. As an optional additional step, the researcher can evaluate different model specifications based on theory and research settings to confirm that the model performs better than the other empirical models. We discuss these in detail in the upcoming paragraphs.

(1) Examine the model for collinearity. First, it is important to ensure the absence of high correlations among the constructs, as these will create methodological and interpretation issues. As with multiple regression, the PLS-SEM algorithm has difficulty estimating models with two predictor constructs that are almost the same (highly correlated) in terms of meaning and construct scores. Hence, we need to ensure the correlations between the constructs are not too high (see step (2) under *formative assessment* above for guidelines when using VIF). If collinearity is present in the model, the researcher can opt to create a second-order construct that is conceptually supported by theory (Sarstedt et al., 2019).
(2) Evaluate the size and significance of the path coefficients. The path coefficients in the inner model are standardized values. In education and L2 research, it can be suggested that path coefficients (β) in the structural model ranging from 0 to .10, .11 to .30, .30 to 50, and > .50 are indicative of weak, modest, moderate, and strong effect sizes. In PLS-SEM, the paths are usually

standardized, but the analysis does not generate the *p*-value by default (Hair et al, 2013; 2022a). Thus, bootstrapping, which uses standard errors, is needed to calculate the *p*-value and bias-corrected confidence intervals. A bootstrap routine with 5,000 (and possibly 10,000) iterations is recommended to achieve stability in the results, and bias-corrected confidence method should be favored.

(3) Assessing the coefficient of determination. Having ensured the model is free of collinearity issues and the paths are meaningful, the next step is assessing the R^2 value in the outcome(s). This estimate reflects the variance in the outcome(s) explained by the predictor constructs. The R^2 measure is commonly used in L2 research. In the PLS-SEM domain, it is referred to as in-sample predictive power because when algorithm solutions are calculated all the sample data is used to calculate the R^2 (Rigdon, 2012). Researchers have emphasised that R^2 values are contextual and should be interpreted with reference to the research domain being examined. For L2 research, we would propose that R^2 values between 0 to .10, .11 to .30, .30 to 50, and > .50 are indicative of weak, modest, moderate, and strong explanatory power, respectively. Note that these rules of thumb should provide general rather than rigid guidance and further investigations are needed because the size of the R^2 may also represent overfit (Hair et al., 2019a). However, model overfit can be controlled for by interpreting the adjusted R^2.

(4) Examine the out-of-sample predictive power. Although R^2 provides useful information about the relationships between all the variables included in the structural and measurement models, it does not assess out-of-sample predictive power as measured by a hold-out sample approach (Hair et al., 2019b; Shmueli et al., 2019). The traditional R^2 metric when interpreted as a measure of model predictive performance is, in reality, an in-sample strength-of-fit – it assesses and explains the relationships between all the variables used to build the model and is thus explanatory in nature (Hair & Sarstedt, 2021). But the R^2 metric is not an indication of the model's out-of-sample prediction performance since the analysis does not predict sample data not included in the initial calculation of the model solution. To obtain out-of-sample prediction metrics researchers must apply a hold-out sample approach (Hair et al., 2022).

Recent advancements in PLS-SEM have introduced the PLSpredict procedure, which entails following a set of steps to calculate out-of-sample prediction metrics; thus, helping in evaluating external validity for similar research design contexts. The PLSpredict procedures incorporate model assessments based on an initial training sample (randomly drawn separate sub-sample of the total sample) and estimate the predictive power of the model on a second hold-out sample of data–other than that used in calculating the initial PLS-SEM solution. The other dataset is referred to as a hold-out sample and that second dataset can be used to calculate an out-of-sample prediction measure (Hair et al., 2019a; Shmueli et al., 2016; Hair & Sarstedt, 2021). Specifically, the PLSpredict procedure generates k-fold cross-validation, where k refers to the number of subgroups into which PLSpredict divides the data. Thus, the total data set is split into the desired number of equally sized data subsets (i.e., k). Shmueli et al. (2019) suggest splitting the sample into 10 subgroups (i.e., k = 10) to maintain the accuracy of the PLSpredict results. If the research involves a smaller total sample size, a smaller number of subgroups should be selected (e.g., k = 4 or 5). The procedure then combines the k – 1 subgroups (i.e., nine subgroups) into one training sample used to predict the excluded tenth subgroup (hold-out sample). In short, the combined k – 1 subgroups predict the single hold-out sample and the resulting metrics measure the out-of-sample prediction. PLSpredict provides several metrics that can be used to compare the predictive performance of out-of-sample data. Hair et al. (2019) suggest that researchers compare the results of the root mean squared error (RMSE) in the two PLSpredict analyses: predictions by the actual observations (PLS-SEM) and the naïve linear regression model (LM), which consists of the indicator means from the analysis sample (Hair et al., 2019a). Sufficient predictive power can be established when the PLS model produces lower prediction error, compared to the naïve LM benchmark (for a detailed explanation and statistical considerations, see Shmueli et al., 2019). The following guidelines should be used to establish predictive power in PLSpredict:

- If the PLS-SEM results show greater prediction errors in RMSE (i.e., higher values than the naïve LM benchmark) for all indicators, this implies the model lacks out-of-sample predictive power.
- If the PLS-SEM results show higher errors in RMSE (i.e., higher values than the naïve LM benchmark) for the majority of the indicators, this implies the model has low predictive power.
- If the PLS-SEM results show higher errors in RMSE (i.e., higher values than the naïve LM benchmark) for the minority of the indicators, this implies the model has medium predictive power.
- If the PLS-SEM results show lower errors in RMSE (i.e. lower values than the naïve LM benchmark) for all indicators, this implies the model has strong predictive power.

It is useful to indicate that achieving favorable out-of-sample predictive power through PLSpredict provides guidance to L2 researchers in evaluating the external validity of their structural model by showing the results can be extended to a similar context. To obtain results for the structural assessment measures described earlier, we use the following code in SEMinR:

```
model <- estimate_pls(data = PLS_SEM_in_L2_2,
measurement_model = mm,
structural_model = sm
)
summary(model)
model_summary <-summary(model)
model_summary$validity$vif_items
model_summary$validity$htmt
model_summary$vif_antecedents
boot_estimates <- bootstrap_model(model, nboot = 5000, cores = 2)
```

```
summary(boot_estimates)
```

To obtain results for PLSpredict, the user needs to use this additional code:

```
predict_simple_model <- predict_pls(
model = model,technique = predict_DA,
noFolds = 10,reps = 10)
summary(predict_simple_model)
```

Using the codes provided above, the user should get the results shown in Appendix A. Note that your bootstrap results will be different from what is reported in this paper.

Should model fit indices be reported in PLS-SEM?

The description thus far did not include discussion about model fit indices, and the reader may wonder where and when to report them. Basically, PLS-SEM was developed as a predictive approach similar to multiple regression analysis. Given this nature, model fit indices were not established for PLS-SEM, but evaluations of their performance for rejecting misfitted models have been carried out in recent years. Among these metrics are the goodness-of-fit index (GoF), standardized root mean square residual (SRMR), the Euclidean distance (d_L), and the geodesic distance (d_G). Proponents of model-fit-indices are skeptical about setting cut-off values to determine the structural model misfit because of the instability of their performance in detecting misspecification in the context of PLS-SEM. Apart from this discussion, the L2 user of PLS-SEM should focus more on establishing the predictive validity of their structural model than on theory testing (Alamer, 2022d; Hair et al., 2022a; Sparks & Alamer, 2022). Hence, model fit indices can be useful only for those investigations that follow a purely confirmatory objective within PLS context.

In contrast to psychometric studies that involve pure investigation of the measurement model, L2 researchers apply structural/path models to understand explanatory, prediction, and path directionality. Accordingly, they should be more concerned with the performance of the directionality and predictive capability of their hypothesized model (Alamer & Alrabai, 2022; Alamer & Lee, 2021). Therefore, pursuing model fit indices when the objective is to understand the effects of variables on other variables has little value. Critically, in some cases, model fit indices can perform against predictive assessment (Hair, 2021). Thus, although the researcher has access to these indices through SEMinR to report them, the user should not reject the PLS structural model based on only the model fit indices (Hair & Sarstedt, 2021). Instead, the researcher should compare the performance of different structural model specifications that are justified by theory and logic to support the selection of the structural model. For example, the user can compare the performance of a full mediation model with a model that have all direct paths to the outcome estimated. In this case, the researcher can check how the BIC indices behave in the two (or more) model configurations (Alamer, 2022c; Hair et al., 2021). As Hair et al. (2019) noted, BIC achieves “a sound tradeoff between model fit and predictive power in the estimation of PLS path models” (p. 8). BIC (among other model criterion indices) can be obtained by requesting the following code in SEMinR: model_summary$it_criteria.

The results of the study example

The detailed results of the analysis of the example study are shown in Appendix A. Starting with the outer model, except for item ‘CI6’, all items appear to be statistically significant and sufficiently large in size (i.e., CI6 loading was .58, p < .001). We retain this item because removing it reduces the AVE and reliability measures. Since the factor loadings of the items are above .70, with only one exception, we can assess indicator reliability. Cronbach’s (α) and CR were all above .70 and less than .95, indicating reliable constructs. In addition, the AVEs were all above the cut-off value of .50, thus establishing the convergent validity of the reflective constructs. With respect to the HTMT ratios for discriminant validity of the constructs, all values were below the conservative .85 cut-off value showing discriminant validity between the constructs.

The formative outcome variable, L2 academic achievement, was evaluated next. The convergent validity (redundancy) evaluates the extent to which the construct predicts a single measure of achievement—self-perception of language proficiency. The results show the path coefficient was .61 (p < .001), suggesting sufficient convergent validity with the proxy benchmark construct. In addition, the VIF values were < 3 so high collinearity is not an issue. Finally, the weights of the two items are significant (p < .001) with w = .90 and .22 for English GPA and teacher’s evaluation, respectively. After establishing the validity of the formative measurement model, we now assess the inner structural model.

As the VIF results indicate, the model did not show signs of collinearity among the constructs, as the highest VIF value was below 3. Next, we examine the path coefficients of our theoretical structural model, first exploring the relationships between the three exogenous antecedents of CI (full results have been provided in Appendix A). We found that need for autonomy was the only significant positive relationship with CI (β = .18, CI 95% [.02, .34]), revealing modest effect size. Contrary to expectations, the sense of competence (β = –.30, CI 95% [–.43, –.15]) and relatedness (β = –.15, CI 95% [–.32, –.01]) were significant and negatively related to CI, with modest effect sizes.

We then examined the relationships between the three exogenous antecedents of PE. In contrast to CI, the antecedents psychological needs of autonomy (β = .15, CI 95% [.02, .34]; modest effect size), competence (β = .45, CI 95% [.33 .55]; moderate effect size), and relatedness (β = .14, CI 95% [.03, .25]; modest effect size) all exhibited positive and significant relationships with PE. Finally, when examining the direct relationships from the three exogenous constructs to the ultimate outcome construct, L2 academic achievement, we observed that two of the basic psychological needs constructs, autonomy and relatedness, were positively associated with the outcome (β = .26, CI 95% [.17, .37] and β = .25, CI 95% [.16, .37], showing modest effect sizes respectively), and that only PE from grit was able to predict scores on the outcome (β = .34, CI 95% [.17, .49]; moderate effect size).

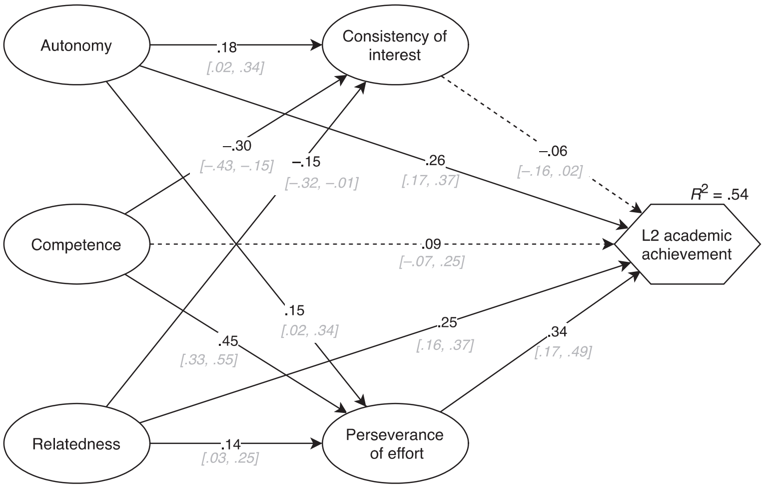

Fig. 3. The structural model.
Note. dashed lines indicate nonsignificant paths; values in brackets are confidence interval (CI) 95%.

Next, we focus on the explained variance in the ultimate outcome variable, L2 academic achievement. Note the outcome variable has been substantially explained by the predictor constructs (R^2 = .54). Some researchers have recommended using the adjusted R^2 (Larson-Hall, 2015; Plonsky, 2015), which yielded a value of .52. Recall from our earlier discussion, however, that more recent research clarifies that R^2 is a measure of in-sample prediction and thus represents an explanatory metric (Hair et al., 2022). Thus, we now evaluate the PLSpredict metrics to determine the out-of-sample prediction power and gather evidence for external validity for similar research design contexts. To do so, we examine the prediction metrics and compare the RMSE of each indicator of achievement (i.e., ach1 and ach2) in PLS-SEM (labelled 'PLS out-of-sample metrics' in the table) with LM (labelled 'LM out-of-sample metrics' in the table). Note the table only lists the indicators that are exclusively directly linked to the outcome (hence, autonomy, competence, and relatedness constructs are excluded).

The analysis shows that all indicators have lower prediction errors in terms of RMSE in the indicators of the outcome variable (see section "examine the predictive power" above). Thus, our PLS model predicts the outcome better than the naïve LM benchmark, indicating strong predictive power and acceptable claim of external validity for similar contexts. As an additional step, we compare the theoretical model with alternative competing models with different specifications. The model with lower values for BIC in the outcome variable has lower measurement errors, and thus should be retained (Alamer, 2022c; Hair et al., 2019). In our case, we tested a competing model where the paths from autonomy, competence and relatedness are removed (i.e., full mediation model). Our results show that the BIC in the original model was −132.17 compared to the competing model –84.325. Accordingly, the original model is retained over the competing model. Thus, the findings show the constructs of autonomy, competence, and relatedness (from basic psychological needs theory; Ryan & Deci, 2017) as well as CI and PE (from grit theory; Duckworth et al., 2007) have successfully and substantially predicted language achievement. These results support previous studies (Alamer, 2021b, 2022d; Alamer & Al Sultan, 2022; Duckworth et al., 2007; Ryan & Deci, 2017) and expand upon them by shedding new light on the relationships between the associated constructs and L2 achievement. For instance, the results indicated that for CI to be sustained, the need of autonomy should be prioritized, while for PE all the basic needs seem to be important. In addition, PE seems to predict L2 achievement better than CI, illustrating possible mediation processes and different trajectories leading to achievement.

Summary and tips for reporting the results

There are similarities and differences between PLS-SEM and the commonly used CB-SEM method. Researchers should first postulate their structural models based on the existing literature. Solutions can then be obtained with PLS-SEM by performing two essential steps: (i) assessing the outer model's validity and reliability, and (ii) assessing the inner model's predictive power. An additional recommended step to consider is (iii) examining competing models. In this paper, we have described these stages in detail and provided the codes needed to run the model using the SEMinR software. We also make the data publicly available for readers to start learning PLS-SEM by replicating our findings.

We recommend researchers report their findings in tables and figures, where possible, and avoid copying and pasting the R output in the paper. In this section, we provide an example of how to report the results obtained in Step 3 of our example study in the form of figures and tables, so that readers can easily peruse all the results (see Fig. 3 and Table 1). Fig. 2 presents information about path coefficients, their significance ($p < .05$), confidence interval, R^2 value. A diagram, which is slightly different from the one presented here, can be readily generated in SEMinR package by inserting the following code: plot(model, title = "PLS Model"),or: plot(boot_estimates, title = "Bootstrapped Model"). Note that constructs represented in hexagon denote formative constructs (composite).

The remaining assessment is $PLS_{predict}$, the results of which should be presented in a table (see Table 1 for an example) showing the RMSE for the PLS model vs. the RMSE for the naïve LM benchmark. Also, recent developments in the PLS-SEM suggest the use

Table 1
RMSE values of items in the PLS model and the naïve benchmark (out-of-sample) model

Indicators of the outcome variable	PLS modelRMSE	Naïve benchmark model (LM) RMSE
ach1	**.540**	.589
ach2	**.989**	1.077

Note. Lower error prediction (RMSE) values are in bold.

of the cross-validated predictive ability test (CVPAT; Liengaard et al., 2020, see also Hair, 2021) for model comparisons based on prediction capabilities. However, the current version of SEMinR has not implemented this procedure yet[1].

Recommendations and conclusions

PLS-SEM is a useful approach to estimating structural models in L2 and education research. Considering its features and the research situations that suit its objectives, such as analyzing complex theoretical models, handling non-normal data, achieving statistical power with smaller sample sizes, and focusing on the model's predictive capability, the method can be an attractive, quite useful alternative to its counterpart, CB-SEM. In fact, PLS-SEM is increasingly applied in social sciences research, but has just made its way into L2 and education domains (Alamer, 2022b, 2022d; Elahi Shrivan & Alamer, 2022; Sparks & Alamer, 2022). We hope this introductory paper will help readers to understand, consider, and apply the method to their research where appropriate. Although it might be tempting to compare the results of PLS-SEM with those generated by CB-SEM, this is not a recommended approach (Hair et al., 2022a) simply because each method uses a distinct algorithm to estimate model parameters, meaning that we expect differences in the results. Indeed, and importantly, the structural model used in this paper is technically 'not identified' using the CB-SEM method. Hence, the application of PLS-SEM is a good alternative for estimating structural models that have special requirements such as having formative constructs (composites).

In addition, we acknowledge that there are different views on PLS-SEM, with each camp advocating for different perspectives, terminologies, and guidelines. For example, a group of researchers recommend the use of consistent PLS (PLSc) method instead of the standard PLS when reflective constructs are involved. However, PLSc aims to mimic those of CB-SEM. Furthermore, PLSc is subject to issues that have been noted for CB-SEM including inaccurate results in certain configurations and is arguably unsuitable for out-of-sample prediction purposes (Hair & Sarstedt, 2021; Sarstedt et al., 2016). Thus, it is a mystery why the user would want to employ PLSc since it produces similar results as CB-SEM while CB-SEM is already there and widely applied. Another area that generated a different view on the subject is the reliance on model fit indices to determine the adequacy of structural models in the PLS-SEM context. We have explained this matter in an earlier section, and we conclude that model fit indices can be useful only for those investigations that follow a purely confirmatory objective. Applying PLS-SEM on structural/path models should, instead, follow explanatory plus prediction assessment as we have illustrated throughout the paper. Terminologies such as *emergent variables, artifacts, formative constructs, auxiliary theory, synthesis theory,* and *forged concepts* are also debatable in the context of PLS-SEM. In this paper, we presented the mainstream thinking on PLS-SEM which is led by Hair and associates (2011, 2022). In addition, PLS-SEM has not yet implemented the ability to estimate cross-leadings at the measurement model level as suggested by the ESEM application (Alamer, 2022c, 2021a), a limitation that should be considered.

This paper described the most recently developed guidelines in PLS-SEM research (Hair et al., 2019a; Hair et al., 2020, 2021; Shmueli et al., 2019; Hair et al., 2022a), which replaced some of the previous guidelines (Hair et al., 2011; 2013). For this reason, we encourage L2 and education researchers to apply the most recent guidelines as outlined in this paper. This methodological article is a good starting point for researchers who have not yet applied PLS-SEM, or SEM in general, in the specification and assessment of a structural model. We hope reviewers and editors of education- and L2-related journals review the discussion provided in this paper, so they will be apprised of the advancements that have recently been made in PLS-SEM. In this paper, we explained the basics of how to apply the PLS-SEM method using the free R package (i.e., SEMinR), though paid software alternatives exist, in the hope that readers can try this method at no cost. It should be noted, however, that this article is a brief overview of the method of PLS-SEM, and there are many other advantages and disadvantages to be explored about the method. Finally, we believe researchers in the field will be able to use PLS-SEM more frequently because the advantages the method provides are well suited to the research contexts they often encounter, as well as to the characteristics of the quantitative data they typically use.

Declaration of Competing Interest

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

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.rmal.2022.100027.

[1] Researchers interested in applying CVPAT can check Liengaard et al. (2021). Also, GitHub repository provides a description of this example and its R code: https://github.com/ECONshare/CVPAT.

Appendix A

```
Results from package seminr (2.2.1)

Path Coefficients:
          PE     CI Achievement
R^2    0.365  0.126       0.536
AdjR^2 0.356  0.113       0.524
AUT    0.149  0.181       0.263
COM    0.447 -0.301       0.092
REL    0.142 -0.151       0.250
CI         .      .      -0.061
PE         .      .       0.344

Reliability:
Alpha, rhoC, and rhoA should exceed 0.7 while AVE should exceed 0.5

            alpha  rhoC   AVE  rhoA
AUT         0.743 0.838 0.566 0.760
COM         0.810 0.875 0.637 0.816
REL         0.781 0.858 0.602 0.787
CI          0.837 0.874 0.541 0.873
PE          0.855 0.893 0.582 0.861
Achievement 0.510 0.749 0.619 1.000

<HTMT>
              AUT   COM   REL    CI    PE Achievement
AUT             .     .     .     .     .           .
COM         0.496     .     .     .     .           .
REL         0.463 0.552     .     .     .           .
CI          0.176 0.328 0.253     .     .           .
PE          0.448 0.675 0.467 0.232     .           .
Achievement 0.749 0.775 0.758 0.253 0.825           .

<VIF>
PE :
  AUT   COM   REL
1.230 1.344 1.315

CI :
  AUT   COM   REL
1.230 1.344 1.315

Achievement :
  AUT   COM   REL    CI    PE
1.301 1.769 1.374 1.145 1.575


Bootstrapping model using seminr...
SEMinR Model successfully bootstrapped

Results from Bootstrap resamples:  5000

Bootstrapped Structural Paths:
                     Original Est. Bootstrap Mean Bootstrap SD T Stat. 2.5% CI 97.5% CI
AUT  ->  CI                  0.181          0.173        0.085   2.133   0.000    0.326
AUT  ->  PE                  0.149          0.153        0.058   2.596   0.043    0.267
AUT  ->  Achievement         0.263          0.268        0.051   5.116   0.161    0.366
COM  ->  CI                 -0.301         -0.306        0.071  -4.250  -0.427   -0.155
COM  ->  PE                  0.447          0.443        0.060   7.435   0.315    0.548
COM  ->  Achievement         0.092          0.082        0.082   1.125  -0.072    0.239
REL  ->  CI                 -0.151         -0.157        0.086  -1.757  -0.337    0.007
REL  ->  PE                  0.142          0.146        0.059   2.430   0.029    0.255
REL  ->  Achievement         0.250          0.260        0.051   4.937   0.161    0.355
```

```
CI  ->  Achievement        -0.061         -0.061        0.048  -1.256  -0.161    0.033
PE  ->  Achievement         0.344          0.339        0.081   4.243   0.186    0.496

Bootstrapped Weights:
                      Original Est. Bootstrap Mean Bootstrap SD T Stat. 2.5% CI 97.5% CI
CI1  ->  CI                   0.225          0.226        0.052   4.308   0.138    0.335
CI2  ->  CI                   0.327          0.318        0.059   5.539   0.226    0.444
CI3  ->  CI                   0.274          0.272        0.042   6.522   0.191    0.359
CI4  ->  CI                   0.255          0.252        0.039   6.530   0.178    0.329
CI5  ->  CI                   0.153          0.153        0.054   2.816   0.033    0.236
CI6  ->  CI                   0.076          0.080        0.068   1.116  -0.067    0.192
PE1  ->  PE                   0.230          0.232        0.026   8.851   0.190    0.289
PE2  ->  PE                   0.188          0.187        0.019  10.099   0.150    0.220
PE3  ->  PE                   0.244          0.243        0.020  12.412   0.212    0.282
PE4  ->  PE                   0.240          0.240        0.018  13.280   0.203    0.275
PE5  ->  PE                   0.188          0.189        0.021   8.813   0.147    0.230
PE6  ->  PE                   0.217          0.216        0.020  10.924   0.173    0.248
aut1  ->  AUT                 0.372          0.370        0.053   6.996   0.275    0.484
aut2  ->  AUT                 0.387          0.390        0.040   9.757   0.325    0.474
aut3  ->  AUT                 0.325          0.326        0.043   7.558   0.238    0.416
aut4  ->  AUT                 0.237          0.235        0.047   5.083   0.139    0.329
com1  ->  COM                 0.280          0.279        0.022  12.518   0.231    0.321
com2  ->  COM                 0.307          0.309        0.030  10.303   0.249    0.368
com3  ->  COM                 0.346          0.347        0.025  13.791   0.306    0.400
com4  ->  COM                 0.318          0.318        0.029  10.983   0.262    0.374
rel1  ->  REL                 0.318          0.319        0.030  10.490   0.260    0.380
rel2  ->  REL                 0.266          0.265        0.037   7.284   0.191    0.332
rel3  ->  REL                 0.370          0.370        0.039   9.472   0.299    0.447
rel4  ->  REL                 0.333          0.334        0.035   9.545   0.268    0.408
ach1  ->  Achievement         0.902          0.903        0.045  19.875   0.813    0.984
ach2  ->  Achievement         0.222          0.215        0.080   2.792   0.055    0.364

Bootstrapped Loadings:
                      Original Est. Bootstrap Mean Bootstrap SD T Stat. 2.5% CI 97.5% CI
CI1  ->  CI                   0.713          0.710        0.054  13.100   0.604    0.809
CI2  ->  CI                   0.814          0.807        0.039  20.784   0.723    0.864
CI3  ->  CI                   0.838          0.832        0.043  19.354   0.760    0.882
CI4  ->  CI                   0.791          0.785        0.046  17.035   0.692    0.849
CI5  ->  CI                   0.639          0.634        0.090   7.072   0.439    0.761
CI6  ->  CI                   0.584          0.580        0.096   6.085   0.395    0.730
PE1  ->  PE                   0.699          0.701        0.040  17.273   0.623    0.774
PE2  ->  PE                   0.745          0.741        0.045  16.479   0.650    0.819
PE3  ->  PE                   0.812          0.812        0.022  36.562   0.766    0.852
PE4  ->  PE                   0.850          0.850        0.030  28.329   0.783    0.897
PE5  ->  PE                   0.703          0.703        0.044  15.940   0.610    0.780
PE6  ->  PE                   0.758          0.754        0.052  14.686   0.634    0.839
aut1  ->  AUT                 0.703          0.695        0.065  10.843   0.547    0.792
aut2  ->  AUT                 0.839          0.837        0.038  22.310   0.757    0.894
aut3  ->  AUT                 0.793          0.789        0.049  16.094   0.673    0.862
aut4  ->  AUT                 0.662          0.654        0.064  10.280   0.502    0.758
com1  ->  COM                 0.749          0.741        0.047  16.014   0.643    0.817
com2  ->  COM                 0.781          0.779        0.036  21.682   0.693    0.839
com3  ->  COM                 0.841          0.840        0.023  35.828   0.790    0.880
com4  ->  COM                 0.818          0.815        0.038  21.764   0.731    0.875
rel1  ->  REL                 0.779          0.776        0.038  20.327   0.687    0.835
rel2  ->  REL                 0.752          0.747        0.045  16.886   0.646    0.824
rel3  ->  REL                 0.777          0.776        0.039  20.162   0.698    0.839
rel4  ->  REL                 0.795          0.795        0.033  24.168   0.722    0.849
ach1  ->  Achievement         0.978          0.976        0.016  62.274   0.938    0.999
ach2  ->  Achievement         0.531          0.521        0.090   5.896   0.330    0.672

Bootstrapped HTMT:
                     Original Est. Bootstrap Mean Bootstrap SD 2.5% CI 97.5% CI
AUT  ->  COM                 0.496          0.502        0.103   0.300    0.696
AUT  ->  REL                 0.463          0.460        0.083   0.294    0.617
AUT  ->  CI                  0.176          0.215        0.045   0.138    0.315
```

```
AUT  ->  PE                 0.448          0.448        0.081  0.274   0.599
AUT  ->  Achievement        0.749          0.755        0.088  0.586   0.933
COM  ->  REL                0.552          0.560        0.106  0.334   0.745
COM  ->  CI                 0.328          0.347        0.071  0.217   0.497
COM  ->  PE                 0.675          0.672        0.062  0.535   0.780
COM  ->  Achievement        0.775          0.780        0.107  0.572   0.993
REL  ->  CI                 0.253          0.276        0.078  0.144   0.448
REL  ->  PE                 0.467          0.469        0.077  0.314   0.622
REL  ->  Achievement        0.758          0.771        0.102  0.554   0.960
CI  ->  PE                  0.232          0.265        0.043  0.195   0.368
CI  ->  Achievement         0.253          0.299        0.107  0.142   0.551
PE  ->  Achievement         0.825          0.830        0.105  0.643   1.056

Bootstrapped Total Paths:
                     Original Est. Bootstrap Mean Bootstrap SD 2.5% CI 97.5% CI
AUT  ->  CI                 0.181          0.172        0.085  0.005   0.326
AUT  ->  PE                 0.149          0.153        0.058  0.043   0.267
AUT  ->  Achievement        0.304          0.311        0.056  0.199   0.420
COM  ->  CI                -0.301         -0.306        0.071 -0.427  -0.155
COM  ->  PE                 0.447          0.443        0.060  0.315   0.548
COM  ->  Achievement        0.264          0.250        0.084  0.081   0.409
REL  ->  CI                -0.151         -0.157        0.086 -0.337   0.007
REL  ->  PE                 0.142          0.146        0.059  0.029   0.255
REL  ->  Achievement        0.308          0.320        0.056  0.211   0.432
CI  ->  Achievement        -0.061         -0.061        0.048 -0.161   0.033
PE  ->  Achievement         0.344          0.339        0.081  0.186   0.496

<PLSpredict>
PLS in-sample metrics:
       PE1   PE2   PE3   PE4   PE5   PE6   CI1   CI2   CI3   CI4   CI5   CI6  ach1  ach2
RMSE 0.996 1.085 0.784 0.860 0.962 0.786 1.051 1.208 1.252 1.221 1.192 1.236 0.540 0.989
MAE  0.822 0.880 0.616 0.658 0.785 0.598 0.880 1.027 1.055 1.031 0.996 1.041 0.426 0.792

PLS out-of-sample metrics:
       PE1   PE2   PE3   PE4   PE5   PE6   CI1   CI2   CI3   CI4   CI5   CI6  ach1  ach2
RMSE 1.009 1.102 0.796 0.870 0.979 0.797 1.065 1.229 1.274 1.238 1.211 1.253 0.569 1.009
MAE  0.834 0.894 0.627 0.663 0.796 0.605 0.890 1.044 1.072 1.046 1.010 1.052 0.445 0.807

LM in-sample metrics:
       PE1   PE2   PE3   PE4   PE5   PE6   CI1   CI2   CI3   CI4   CI5   CI6  ach1  ach2
RMSE 0.823 0.981 0.686 0.763 0.860 0.694 0.972 1.043 1.148 1.109 1.090 1.065 0.496 0.915
MAE  0.647 0.762 0.532 0.583 0.694 0.544 0.781 0.822 0.937 0.897 0.885 0.866 0.389 0.735

LM out-of-sample metrics:
       PE1   PE2   PE3   PE4   PE5   PE6   CI1   CI2   CI3   CI4   CI5   CI6  ach1  ach2
RMSE 0.921 1.153 0.781 0.882 0.992 0.813 1.113 1.187 1.328 1.246 1.237 1.233 0.589 1.077
MAE  0.722 0.890 0.608 0.659 0.777 0.632 0.896 0.933 1.077 1.006 1.000 0.995 0.460 0.863
```

Appendix B

L2-Grit Scale (Alamer, 2021b)

Item	Factor
I work hard towards my language learning goals irrespective of how long they take to achieve.	PE
Even when I can do something more fun, I give language learning tasks my best effort.	PE
I complete my language learning tasks irrespective of how difficult they are.	PE
I am committed to the investment of my best effort in language learning tasks.	PE
Even if I am struggling to learn the language, I keep trying my best.	PE
Once I set a language learning goal, I try to overcome any challenge that arises.	PE
I often set a language learning goal but later choose to pursue a different one.	CI
New ideas and projects sometimes distract me from learning the language.	CI
I become interested in new pursuits other than language learning every few months.	CI
My interest in learning the second language changes every month	CI
I was obsessed with learning the language for a short period of time but lost interest eventually.	CI
I have difficulty in maintaining my focus on language tasks that take a long time to achieve.	CI

Basic Psychological Needs of Second Language Scale (BPN-L2) (Alamer, 2022a)

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.
My English teacher let me freely practise 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.

References

Alamer, A. (2022a). Basic psychological needs, motivational orientations, effort, and vocabulary knowledge: A comprehensive model. *Studies in Second Language Acquisition, 44*(1), 164–184. 10.1017/S027226312100005X.

Alamer, A. (2022b). Basic psychological need satisfaction and continued language learning during a pandemic: A structural equation modelling approach. *Journal for the Psychology of Language Learning, 4*(1). 10.52598/jpll/4/1/1.

Alamer, A. (2022c). Exploratory structural equation modeling (ESEM) and bifactor ESEM for construct validation purposes: Guidelines and applied example. *Research Methods in Applied Linguistics, 1*(1), Article 100005. 10.1016/j.rmal.2022.100005.

Alamer, A. (2022d). Having a single language interest autonomously predicts L2 achievement: Addressing the predictive validity of L2 grit. *System, 108*, Article 102850. 10.1016/j.system.2022.102850.

Alamer, A. (2021a). Construct validation of self-determination theory in second language scale (SDT-L2): The bifactor-exploratory structural equation modeling (ESEM) approach. *Frontiers in Psychology*. 10.3389/fpsyg.2021.732016.

Alamer, A. (2021b). Grit and language learning: Construct validation of grit and its relation to later vocabulary knowledge. *Educational Psychology, 41*(5), 544–562. 10.1080/01443410.2020.1867076.

Alamer, A., & Almulhim, F. (2021). The interrelation between language anxiety and L2 motivation; A mixed methods approach. *Frontiers in Education*. 10.3389/feduc.2021.618655.

Alamer, A., & Alrabai, F. (2022). The causal relationship between learner motivation and language achievement: New dynamic perspective. *Applied Linguistics*. 10.1093/applin/amac035.

Alamer, A., & Al Sultan, H. (2022). The role of basic psychological needs on volunteering, and national responsibility during the COVID-19 pandemic: Results from the context of Saudi Arabia. *Frontiers in Education*. 10.3389/feduc.2022.944048.

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., & 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., & 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.

Chin, W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern Methods for Business Research* (pp. 295–358). Mahwah: Erlbaum.

Chin, W., Cheah, J., Liu, Y., Ting, H., Lim, X., & Cham, T. (2020). Demystifying the role of causal-predictive modeling using partial least squares structural equation modeling in information systems research. *Industrial Management & Data Systems, 120*(12), 2161–2209. 10.1108/IMDS-10-2019-0529.

Duckworth, A., Peterson, C., Matthews, M., & Kelly, D. (2007). Grit: Perseverance and passion for long-term goals. *Journal of personality and social psychology, 92*, 1087–1101. 10.1037/0022-3514.92.6.1087.

Elahi Shrivan, M., & Alamer, A. (2022). Modeling the interplay of EFL learners' basic psychological needs, grit and L2 achievement. *Journal of Multilingual and Multicultural Development*. 10.1080/01434632.2022.2075002.

Hair, J., Black, W., Babin, B., & Anderson, R. (2019). *Multivariate data analysis* (8th ed.). Cengage.

Hair, J., Hult, G., Ringle, C., & Sarstedt, M. (2013). *A primer on partial least squares structural equation modeling (PLS-SEM)* (1st ed.). SAGE.

Hair, J., Hult, G., Ringle, C., & Sarstedt, M. (2022a). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3nd ed.). SAGE.

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2022b). *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R*. New York: Springer.

Hair, J. F., Hult, G. T., Ringle, C. M., Sarstedt, M., & Thiele, K. O. (2017). Mirror, mirror on the wall: A comparative evaluation of composite-based structural equation modeling methods. *Journal of the Academy of Marketing Science, 45*, 616–632.

Hair, J., Ringle, C., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice, 19*(2), 139–152. 10.2753/MTP1069-6679190202.

Hair, J. F., Ringle, C. M., Gudergan, S. P., Fischer, A., Nitzl, C., & Menictas, C. (2018). Partial Least Squares Structural Equation Modeling-based discrete choice modeling: an illustration in modeling retailer choice. *Business Research, 12*(1), 115–142 Springer: German Academic Association for Business ResearchApril.

Hair, J., Risher, J., Sarstedt, M., & Ringle, C. (2019a). When to use and how to report the results of PLS-SEM. *European Business Review, 31*(1), 2–24. 10.1108/EBR-11-2018-0203.

Hair, J. F., Sarstedt, M., & Ringle, C. M. (2019b). Rethinking some of the rethinking of partial least squares. *European Journal of Marketing, 53*(4), 566–584.

Hair, J. F., & Sarstedt, M. (2021). Factors versus composites: Guidelines for choosing the right structural equation modeling method. *Project Management Journal, 50*(6), 619–624.

Hair, J. F., & Sarstedt, M. (2020). Factors versus composites: Guidelines for choosing the right structural equation modeling method. *Project Management Journal, 50*(6), 619–624. 10.1177/8756972819882132.

Hair, J. F., & Sarstedt, M. (2021). Explanation plus prediction – The logical focus of project management research. *Project Management Journal*. Advance online publication. https://doi.org/10.1177%2F8756972821999945

Hair, J. F. (2021). Next generation prediction metrics for composite-based PLS-SEM. *Industrial Management & Data Systems, 121*(1), 5–11. 10.1108/IMDS-08-2020-0505.

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

Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma–exponential methods. *Information Systems Journal, 28*(1), 227–261.

Larson-Hall, J. (2015). *A guide to doing statistics in second language research using SPSS and R* (2nd ed.). Routledge. 10.4324/9781315775661.

Latan, H., & Noonan, R. (2017). *Partial least squares path modeling: Basic concepts, methodological issues and applications*. Springer. 10.1007/978-3-319-64069-3.

Law, L., & Fong, N. (2020). Applying partial least squares structural equation modeling (PLS-SEM) in an investigation of undergraduate students' learning transfer of academic English. *Journal of English for Academic Purposes, 46*, Article 100884. 10.1016/j.jeap.2020.100884.

Liengaard, B., Sharma, P. N., Hult, G. T. M., Jensen, M. B., Sarstedt, M., Hair, J. F., & Ringle, C. M. (2020). Prediction: coveted, yet forsaken? Introducing a cross-validated predictive ability test in partial least squares path modeling. *Decision Sciences*. 10.1111/deci.12445.

Mohammadi, R. R., Saeidi, M., & Ahangari, S. (2020). Self-regulated learning instruction and the relationships among self-regulation, reading comprehension and reading problem solving: PLS-SEM approach. *Cogent Education, 7*(1), Article 1746105. 10.1080/2331186X.2020.1746105.

Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology, 94*(2), 249.

Noels, K. (2013). Learning Japanese; learning English: Promoting motivation through autonomy, competence and relatedness. *Foreign Language Motivation*, 15–34 *Japan*.

Noels, K., Pelletier, L., Clément, R., & Vallerand, R. (2000). Why are you learning a second language? Motivational orientations and self-determination theory. *Language Learning, 50*(1), 57–85. 10.1111/0023-8333.00111.

Oga-Baldwin, Q., Nakata, Y., Parker, P., & Ryan, R. (2017). Motivating young language learners: A longitudinal model of self-determined motivation in elementary school foreign language classes. *Contemporary Educational Psychology, 49*, 140–150. 10.1016/j.cedpsych.2017.01.010.

Ray, S., Danks, N. P., & Velasquez-Estrada, J. M. (2020). Seminr: Domain-specific language for building and estimating structural equation models [computer software]. R package version 1.2.0. Retrieved from: https://cran.r-project.org/web/packages/seminr/index.html

R Core Team. (2020). *R: A language and environment for statistical computing*. https://www.R-project.org

Ringle, C. M., Wende, S., & Becker, J-M. (2015). *SmartPLS 3 [Computer Software]*. Bönningstedt: SmartPLS https://www.smartpls.com/.

Rigdon, E. (2012). rethinking partial least squares path modeling: in praise of simple methods. *Long Range Planning, 45*(5), 341–358. 10.1016/j.lrp.2012.09.010.

Ronkko, M., McIntosh, C. N., & Antonakis, J. (2015). On the adoption of partial least squares in psychological research: Caveat emptor. *Personality and Individual Differences, 87*, 76–84.

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

Sarstedt, M., Hair, J. F., Cheah, J. H., Becker, J. M., & Ringle, C. M. (2019). How to specify, estimate, and validate higher-order constructs in PLS-SEM. *Australasian Marketing Journal (AMJ), 27*(3), 197–211.

Sarstedt, M., Hair, J. F., Nitzl, C., Ringle, C. M., & Howard, M. C. (2020). Beyond a tandem analysis of SEM and PROCESS: Use PLS-SEM for mediation analyses!. *International Journal of Market Research, 62*(3), 288–299. 10.1177/1470785320915686.

Sarstedt, M., Ringle, C., Smith, D., Reams, R., & Hair, J. (2014). Partial least squares structural equation modeling (PLS-SEM): A useful tool for family business researchers. *Journal of Family Business Strategy, 5*(1), 105–115. 10.1016/j.jfbs.2014.01.002.

Shao, K., Elahi Shirvan, M., & Alamer, A. (2022). How accurate is your correlation? Different methods derive different results and different interpretations. *Frontiers in Psychology, 13*, Article 901412. 10.3389/fpsyg.2022.901412.

Shmueli, G., Ray, S., Velasquez Estrada, J., & Chatla, S. (2016). The elephant in the room: Predictive performance of PLS models. *Journal of Business Research, 69*(10), 4552–4564. 10.1016/j.jbusres.2016.03.049.

Shmueli, G., Sarstedt, M., Hair, J., Cheah, J., Ting, H., Vaithilingam, S., & Ringle, C. (2019). Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict. *European Journal of Marketing, 53*(11), 2322–2347. 10.1108/EJM-02-2019-0189.

Sparks, R., & Alamer, A. (2022). Long-term impacts of L1 language skills on L2 anxiety: The mediating role of language aptitude and L2 achievement. *Language Teaching Research*. 10.1177/13621688221104392.

Winter, B. (2019). *Statistics for linguists: An introduction using R*. Wang: Routledge. 10.4324/9781315165547.

Wold, H. (1982). Soft modeling: The basic design and some extensions. *Systems Under Indirect Observation, 2*, 343.