Received: 11 July 2023 | Accepted: 17 January 2024

DOI: 10.1111/bmsp.12336

TUTORIAL

the british psychological society
promoting excellence in psychology

When and how to use set-exploratory structural equation modelling to test structural models: A tutorial using the R package lavaan

Herb Marsh[1,2] | **Abdullah Alamer**[3,4]

[1]Institute of Positive Psychology and Education, Australian Catholic University, Sydney, New South Wales, Australia

[2]Department of Education, University of Oxford, Oxford, UK

[3]Department of English, King Faisal University, Hofuf, Saudi Arabia

[4]The University of New South Wales (UNSW), Sydney, Australia

Correspondence
Abdullah Alamer, King Faisal University, Saudi Arabia.
Email: aa.alamer@kfu.edu.sa

Funding information
Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Grant/Award Number: 5538

Abstract

Exploratory structural equation modelling (ESEM) is an alternative to the well-known method of confirmatory factor analysis (CFA). ESEM is mainly used to assess the quality of measurement models of common factors but can be efficiently extended to test structural models. However, ESEM may not be the best option in some model specifications, especially when structural models are involved, because the full flexibility of ESEM could result in technical difficulties in model estimation. Thus, set-ESEM was developed to accommodate the balance between full-ESEM and CFA. In the present paper, we show examples where set-ESEM should be used rather than full-ESEM. Rather than relying on a simulation study, we provide two applied examples using real data that are included in the OSF repository. Additionally, we provide the code needed to run set-ESEM in the free R package *lavaan* to make the paper practical. Set-ESEM structural models outperform their CFA-based counterparts in terms of goodness of fit and realistic factor correlation, and hence path coefficients in the two empirical examples. In several instances, effects that were non-significant (i.e., attenuated) in the CFA-based structural model become larger and significant in the set-ESEM structural model, suggesting that set-ESEM models may generate more accurate model parameters and, hence, lower Type II error rate.

KEYWORDS

basic psychological needs, confirmatory factor analysis (CFA), exploratory structural equation modelling (ESEM), motivation, set-exploratory structural equation modelling (set-ESEM), structural equation modelling (SEM)

1 | INTRODUCTION

Exploratory structural equation modelling (ESEM) is a statistical technique that combines the advantages of both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Developed by Asparouhov and Muthén (2009) and further elaborated by Marsh et al. (2009, 2014), ESEM allows the examination of complex factor structures while also allowing for model testing and evaluation. ESEM is particularly more useful than the traditional CFA when items have multiple sources of variance (Morin et al., 2013), as it often provides better goodness of fit, less inflated factor correlations, and meaningful cross-loadings. Researchers have continuously reported that ESEM outperforms CFA across different psychological fields, such as clinical, health, industrial psychology (Marsh et al., 2014), and educational psychology (Alamer, Al Khateeb, & Jeno, 2023; Alamer & Marsh, 2022; Alamer, Morin, et al., 2023; Guay et al., 2015; Kruk et al., 2023). However, some empirical situations require restrictions in the fully relaxed ESEM model. This leads to the recently developed *set-ESEM* (Marsh et al., 2020). If ESEM combines EFA and CFA in one analytical framework, set-ESEM combines ESEM and CFA in one balanced framework. Although set-ESEM was introduced in previous work, in this methodological paper, we elaborate on this technique by assessing its performance in *structural models* compared to CFA. In this tutorial paper, we provide a step-by-step application of set-ESEM through a free R statistical package, *lavaan* (Rosseel, 2012), using two empirical studies that are relevant to analysts in the field of educational psychology in general and second language (L2) learning in particular. Although readily available in Mplus software, we think that readers would benefit from using *lavaan* as it is free of charge and has recently supported estimating ESEM (for versions 0.6-13 and above). To make this tutorial paper practical we also share the data used in the illustrative examples to help the reader train and replicate our analysis.

2 | EFA, CFA, ESEM, AND SET-ESEM: WHAT IS THEIR STORY?

The origins of EFA date back to the seminal works of Spearman (1904) and Thurstone (1935, 1947). EFA was not originally referred to as 'exploratory' but only described as 'factor analysis' (Morin, 2023). It was only when CFA came into the picture that the label 'exploratory' was attached to EFA. The arrival of CFA appeared unintentionally to attenuate the role of EFA in examining established factor structures (Alamer, Morin, et al., 2023; Swami et al., 2023). What makes CFA appealing to researchers is that it draws on the feature of SEM. For example, by using CFA, analysts can access goodness-of-fit indices, deal with missing data through advanced methods (such as full-information maximum likelihood), compare the fit of competing models (e.g., via the Bayesian information criterion (BIC)), allowing uniquenesses (i.e., residuals) to be correlated when theory or technical requirements allow for their inclusion, and beyond that, immediately respecify the CFA model as a structural model (i.e., estimating path coefficients between the factors instead of correlation). With these benefits missing from EFA, researchers have relied more on CFA when they want to study a priori hypothesized models. As such, EFA became more relevant when analysts wanted an early impression of the relationship between the items and presumed factors, which could then be tested with CFA for confirmatory purposes. However, CFA rarely supports what EFA suggests, pushing researchers to apply modifications to their model to satisfy model fit indices in the clusters model of confirmatory factor analysis (ICM-CFA; Alamer, 2021a; Kruk et al., 2023; Morin et al., 2013). Apart from that, a key characteristic of ICM-CFA (which is also a drawback) is that it typically assumes that items only load on *one factor*. That is, cross-loadings across other relevant factors in the model are specified to equal zero.

In fact, in many simulation studies, the true population model is specified to have a few cross-loadings and tested with an intentionally misspecified model that does not include these cross-loadings (Hu & Bentler, 1999). Nevertheless, in many research situations, the measurement model has items that load on their presumed factor which should meaningfully be allowed to load on other conceptually relevant factors. Consider, for example, the factor structure of the dualistic model of passion (Vallerand, 2015; see also Alamer & Marsh, 2022) which involves two factors, *obsessive passion* and *harmonious passion*. As the theory postulates, the two factors share similar conceptual meanings. Hence, they are not distinct entities (Horwood et al., 2021). In such a case, specifying a CFA where the items of *obsessive passion* load only on their presumed factor (i.e., the obsessive passion factor) is an overly restrictive approach. Empirical research constantly reflects this dilemma. For example, in almost all empirical research, a CFA solution for the dualistic model of passion results in specific issues and is often not empirically supported by the analysis (see Alamer & Marsh, 2022, for a discussion).

To compensate for the inappropriateness of fit of the CFA, researchers apply ex post facto strategies such as removing several items (or even a whole factor from the measurement model), correlating several error terms, or even proposing an alternative measurement model not originally postulated by the theory. Apparently, heavy reliance on these ex post facto strategies is non-theoretical and mainly seen as dubious and misleading (Alamer, Morin, et al., 2023; Faraci et al., 2023; Guay et al., 2015; Marsh et al., 2014; Morin et al., 2013; Swami et al., 2023; Van Zyl & ten Klooster, 2022). Even if the CFA fits the data (commonly after post hoc modifications), the factor correlations are often inflated, which prevents supporting the discriminant validity between the factors. Shao et al. (2022) showed an applied example and found that the correlation can increase from $r = .51$ in the EFA/ESEM to $r = .82$ in the CFA ($\Delta r = +.31$). They indicated that some research papers obtained a correlation of .90 when relying on CFA. Critically, with such an inflated factor correlation in the CFA, the discriminant validity of the constructs is far from being supported. Most importantly, failure to establish discriminant validity can have determinantal implications when the measurement model is used in a structural model to evaluate the effects of these factors on plausible outcomes. Statistically, collinearity can attenuate the path coefficients and distract from valid structural relationships (Morin, 2023; Shao et al., 2022; Tabachnick & Fidell, 2023). We show a real example of this critical issue in the present study, despite the CFA model fitting the data.

So, if CFA carries some limitations, what can we do? Going back to the standard EFA might not be the best decision due to its limited functionality, although cross-loadings in the EFA are naturalistically appropriate. Similarly, persisting with CFA means that we might end up with an estimated model that suffers from empirical and/or theoretical concerns, although it comes with SEM features. Fortunately, analysts have found a way to integrate the benefits of SEM into the EFA. This development is referred to as ESEM (Asparouhov & Muthén, 2009). ESEM allows EFA to share the advantages that CFA gains from SEM. Thus, ESEM simultaneously permits all cross-loadings among conceptually relevant factors (which is an EFA-derived property) and builds on SEM functionality such as computation of goodness of fit, assessing measurement invariance, dealing with missing values through advanced methods, incorporating covariates in the model, and correlating residuals if needed (which are CFA-derived properties) (Alamer, 2022; Marsh et al., 2009). Figure 1 shows a visual difference between CFA, ESEM, and set-ESEM according to a four-factor measurement model. Note that the dashed lines in Figure 1 refer to non-target loadings, which should be constrained to be as close to zero as possible to achieve the confirmatory approach of ESEM. This can be done either automatically or manually. The automatic method involves using *geomin rotation*, which is a mechanical procedure. An epsilon value of .50 is suggested to reduce the cross-loading sizes. In contrast, the preferred method is the manual method which involves targeting the size of cross-loadings by using *target rotation*. This is a non-mechanical procedure as it involves targeting the cross-loadings to be as close to zero as possible, although they might not necessarily become zero (Morin, 2023).

A recent meta-analytic review (Gegenfurtner, 2022) has illustrated the utility of ESEM in testing multidimensional latent constructs. Reviewing 158 studies ($k = 308$, $N = 778{,}624$), Gegenfurtner (2022) found that ESEM is superior to CFA in different perspectives, as evidenced by better goodness of fit and support for the discriminant validity of the factors. In addition, Van Zyl and ten Klooster (2022) illustrated practical guidelines and developed a useful online tool that helps researchers generate Mplus syntax for ESEM models. Similarly, Morin (2023) provided an excellent overview of and guidelines for

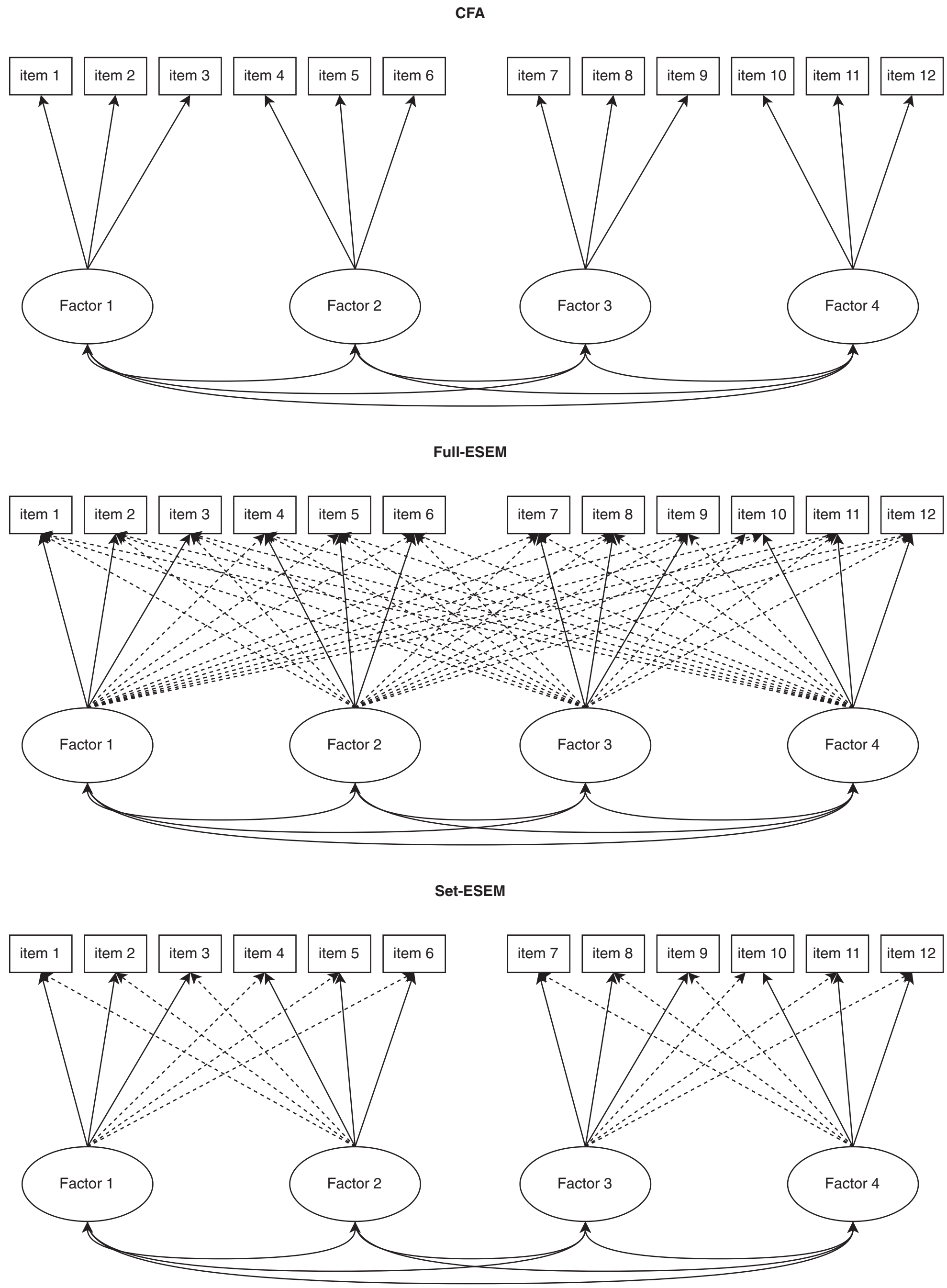

FIGURE 1 A juxtaposition of CFA, full-ESEM, and set-ESEM. *Note*: Dashed lines indicate non-target cross-loadings.

implementing ESEM, noting that 'ESEM has launched a whole new era of statistical research, which has made it possible to realize that by relegating EFA to the role of second-class citizen, we were very close to throwing out the baby with the bathwater' (p. 522). Nevertheless, some would criticize ESEM as being less parsimonious than CFA when providing a better fit. It should be noted, however, that some fit indices, such as the Tucker–Lewis index (TLI) and root mean square error of approximation

(RMSEA), control for parsimony (Alamer, 2022; Morin et al., 2013). If the added parameters in ESEM are less favourable, these indices should be worse in ESEM.

In the L2 domain, ESEM becomes more accepted and expected by applied researchers. The first study in the L2 domain that applied ESEM is perhaps Alamer (2021a), as the study successfully examined the factor structure of the self-determination theory in second language (SDT-L2) scale via ESEM (the scale will be discussed later). Since then, subsequent research in the L2 field has consistently found that relying on CFA and ignoring ESEM can result in detrimental effects on the understanding of the measurement model (Alamer, 2022; Alamer & Marsh, 2022; Alamer, Morin, et al., 2023; Faraci et al., 2023; Kruk et al., 2023; Liu et al., 2022; Shao et al., 2022). It is now widely accepted that ESEM enables researchers to accurately depict the underlying factors, their correlations, and their regressions by utilizing all available data at the indicator level. These studies have pointed out that construct validation appears to be impractical when using CFA because the model suggested by EFA could not be replicated in the CFA.

2.1 | Situations where set-ESEM should be used over full-ESEM

Nevertheless, there are situations where full-ESEM might not be the best choice for the analysis. This can occur when a set of factors and items should align and separate from irrelevant sets of factors and items. One could have mini-sets of ESEM modelled within one comprehensive model in such instances. In such a situation, one speaks of set-ESEM (Marsh et al., 2020). Set-ESEM is a recent development in factor analysis aiming to achieve an optimal compromise between CFA and full-ESEM about the model fit indices, parsimony, confirmatory, and a well-defined measurement model. In some situations, set-ESEM can be relevant and possibly more appropriate than full-ESEM. Here we describe two possible situations with real data and empirical analyses.

The first relates to the situation in which a researcher collects data on items that belong to constructs that are conceptually distinct or belong to different theories. For example, consider a data set that contains items measuring *autonomy*, *competence*, and *relatedness* from the three basic psychological needs (BPN; Ryan & Deci, 2017) using the BPN-L2 scale (Alamer, 2022), and two constructs, *perseverance of effort* and *consistency of interest* from grit theory (Duckworth et al., 2007) using the L2-grit scale (Alamer, 2021b). It might be theoretically inappropriate to estimate cross-loadings between the items of grit on the items of the three BPN, and vice versa. This is because each theory posits different functions for its constructs and their items. For example, the BPN factors are theorized as socially contextual, while grit is posited to be a stable personality trait. In such cases, the analyst may want to consider using two ESEM blocks (or sets), one set that allows cross-loadings between the three BPN but not between the items of grit and another set that allows cross-loadings between the two grit factors but not between the items of the three BPN. This approach achieves a more parsimonious model for theory examination and empirical balance for model testing.

The second situation that necessitates using set-ESEM relates to the setting in which a researcher collects data from theoretically relevant constructs from more than one time point. In this case, cross-loadings should only be allowed to be estimated among the nested constructs at the same time point. For instance, consider a data set containing harmonious and obsessive passion, and autonomy measured at two time points. These three constructs are conceptually relevant such that trivial cross-loadings are reasonable. Nevertheless, allowing cross-loadings from items at time 1 to cross-load on time 2 constructs is theoretically and empirically unsuitable. This might result in confounding effects and is technically undesirable. Apart from that, it is important to note that it is routinely posited in longitudinal SEM analyses that residuals of the same item over time are correlated (Marsh & Hau, 1996). Applying set-ESEM in such a case makes it possible to maintain the flexibility of ESEM as well as the rigour and parsimony of CFA.

In the following sections, we report on two studies to showcase the two scenarios and show why set-ESEM can be a better alternative to the restrictive system of CFA. For the reliability indices it

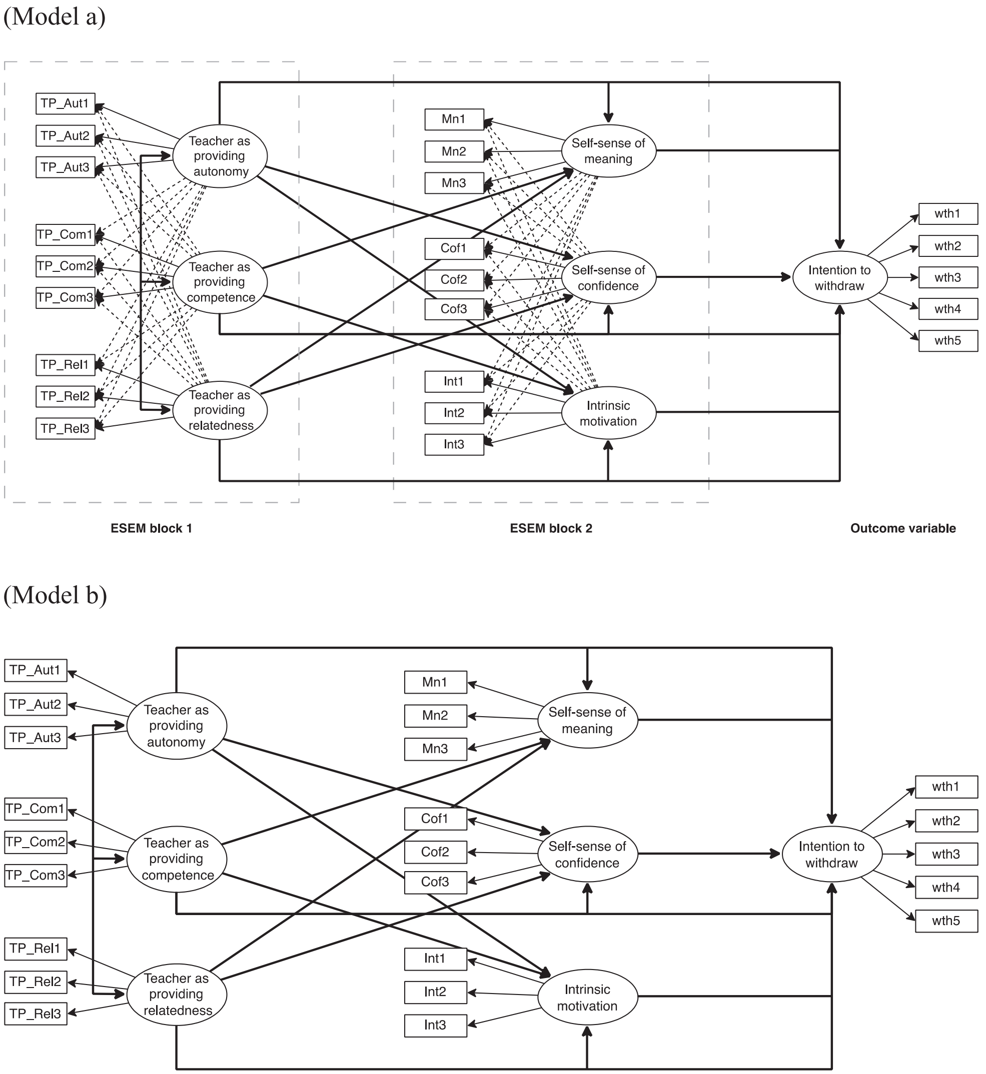

FIGURE 2 Set-ESEM structural model (model A) and CFA-based structural model (model B).

is suggested that researchers report the omega (ω) coefficient of model-based composite reliability. Since this is a methodological demonstration, we focus more on the analyses than on theoretical discussion.

3 | EMPIRICAL EXAMPLES

3.1 | Participants in the two studies

In our two examples, Saudi participants who learn English as an L2 at a public Saudi university participated through an online questionnaire. In the first study 269 students were involved, while in the second study 389 students were involved. All male and female students were between 18 and 20 years of age ($M = 18.5$) and spoke Arabic as their mother tongue. All students had the freedom to ignore the online invitation or withdraw from participating while filling out the questionnaire by simply closing the webpage. The review committee at the university approved the data collection of both research studies. The corresponding author was responsible for data collection in the two studies.

3.2 | Study 1: Using set-ESEM to assess different sets of theories simultaneously

For demonstration purposes, Study 1 reports on the first scenario where the researcher may have different constructs drawn from different theories or levels. In our case, we have two blocks of theory. The first is the three constructs drawn from the BPN theory: autonomy, competence, and relatedness (see Noels, 2023; Ryan & Deci, 2017). Specifically, the measure was students' perceptions about the English teacher as providing the three BPN factors. The second block represents different outcomes that BPN theory suggests, including self-sense of meaning, self-sense of confidence, and intrinsic motivation. Research suggests that when students perceive their teacher as providing healthy BPN, one can expect more intrinsic motivation, sense of meaning, and confidence (Alamer, 2022; Alamer & Al Khateeb, 2023; Alamer, Al Khateeb, & Jeno, 2023; Guay et al., 2015; Noels, 2023). In this study, the corresponding author asked participants to indicate their perception of their English teacher as providing these three BPN factors, while for sense of meaning, confidence, and intrinsic motivation, participants were asked to indicate their self-perception. This domain difference justifies *independent* item cross-loadings. That is, the cross-loadings should not be allowed between other variables that target self-sense (i.e., self-sense of meaning, confidence, and intrinsic motivation) due to the distinct conceptual meaning. It can be conceptually difficult to justify cross-loadings between constructs that ask about the teacher and those related to self-sense of psychological variables. Finally, the intention to withdraw from the course is used as the outcome variable in the model. The two alternative models are presented in Figure 2.

3.2.1 | Measures

We used the BPN-L2 scale (Alamer, 2022), with three items for each construct. Example items are 'My teacher allows us to decide what language learning tasks to do' (autonomy; $\omega = .75$), 'My teacher tells us that we are capable of learning English' (competence; $\omega = .75$), and 'My English teacher is friendly and cordial with us' (relatedness; $\omega = .91$). For intrinsic motivation, three items from the SDT-L2 scale (Alamer, 2022) were adopted. An example item is 'I learn English because I enjoy it' ($\omega = .91$). The three items for self-sense of confidence and three items for meaning were adopted from Dörnyei and Ushioda (2021) as well as Dörnyei and Ryan (2015). An example item for self-confidence is 'I believe in my abilities to do well in the course' ($\omega = .74$) and for meaning 'I know why I have enrolled in this course' ($\omega = .91$). Five items were adopted from Lounsbury et al. (2004) to measure intention to withdraw. An example item is 'I do not plan to continue studying in this major' ($\omega = .90$). The measures are based on a five-point Likert response format ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

3.2.2 | The code needed in lavaan to run the set-ESEM model

The R package *lavaan* (version 0.6-13 and above) was used to estimate the ESEM block(s) in the two studies. Relative to specifying a CFA model, an ESEM block can be specified using the following code[1]:

[1]The user should always double-check on the rotated factor loadings in the output to make sure that the 'meaning' of the factors coincides with the meaning implied by their given names.

```
efa("teacher")*
```

Note that `teacher` is the name we have chosen to give to the first ESEM block in Figure 2 (model A), although we are free to choose any label for this ESEM block. It could be the only ESEM block needed, but in our case, we have two sets of ESEM. Thus, one can specify its factors and items in one instance after indicating that this is the first block. There are two ways to specify the measurement model of ESEM in *lavaan*: the long and short formats. In the short format, the factors in the first ESEM are named in one instance as follows:

```
efa("teacher")*Teacher_autonomy +
efa("teacher")*Teacher_competence +
efa("teacher")*Teacher_relatedness
```

After inserting the names of the factors, one can insert their items by using =~ notation followed by the items. Thus, the first ESEM block can defined as follows:

```
efa("teacher")*Teacher_autonomy +
efa("teacher")*Teacher_competence +
efa("teacher")*Teacher_relatedness =~ T_autonomy1 + T_autonomy2 +
T_autonomy3 + T_competence1 + T_competence2 + T_competence3 +
T_relatedness1 + T_relatedness2 + T_relatedness3
```

In the long format, each factor is defined separately. A major advantage of the long format is that it provides flexibility; for example, labels (model constraints) could be added in specific model parameters. We give an example relating to model constraints in Study 2.

```
efa("teacher")*Teacher_autonomy =~ T_autonomy1 + T_autonomy2 +
T_autonomy3 + T_competence1 + T_competence2 + T_competence3 +
T_relatedness1 + T_relatedness2 + T_relatedness3

efa("teacher")*Teacher_competence =~ T_autonomy1 + T_autonomy2 +
T_autonomy3 + T_competence1 + T_competence2 + T_competence3 +
T_relatedness1 + T_relatedness2 + T_relatedness3

efa("teacher")*Teacher_relatedness =~ T_autonomy1 + T_autonomy2 +
T_autonomy3 + T_competence1 + T_competence2 + T_competence3 +
T_relatedness1 + T_relatedness2 + T_relatedness3
```

After defining the first ESEM block, the user can specify the second ESEM block in the same way. We gave it another name, `self`, as shown below:

```
efa("self")*
```

Then the user can specify the second ESEM block factors and items. Here is the annotated code for the short format:

```
efa("self")*Self_Meaning +
efa("self")*Self_Confidence +
efa("self")*Intrinsic_Motivation =~ S_meaning1 + S_meaning2 +
S_meaning3 + S_confidence1 + S_confidence2 + S_confidence3 +
S_Intrinsic1 + S_Intrinsic2 + S_Intrinsic3
```

TABLE 1 Model fit indices for the two structural models.

Model	χ^2	*df*	RMSEA	RMSEA 90% confidence interval	CFI	TLI
CFA-based model	419.040*	209	.064	(.055, .073)	.942	.930
Set-ESEM-based model	367.121*	185	.063	(.053, .072)	.950	.932

*Significant at $p < .01$.

TABLE 2 Path coefficients in the two models.

		CFA-based results		Set-ESEM results	
Dependent	Predictor	*b**	*p*-Value	*b**	*p*-Value
`Self_Meaning`	`Teacher_autonomy`	−.16	.38	−.02	.83
`Self_Meaning`	`Teacher_competence`	.31	.04	.23	.02
`Self_Meaning`	`Teacher_relatedness`	.28	.05	.28	<.01
`Self_Confidence`	`Teacher_autonomy`	−.08	.61	−.11	.29
`Self_Confidence`	`Teacher_competence`	.41	<.01	.22	.04
`Self_Confidence`	`Teacher_relatedness`	−.38	.01	−.31	.02
`Intrinsic_motivation`	`Teacher_autonomy`	.55	<.01	.46	<.01
`Intrinsic_motivation`	**`Teacher_competence`**	**.08**	**.51**	**.19**	**.03**
`Intrinsic_motivation`	**`Teacher_relatedness`**	**.21**	**.11**	**.31**	**<.01**
`Intent_to_Quit`	`Self_Meaning`	−.13	.08	−.12	.11
`Intent_to_Quit`	`Self_Confidence`	.04	.55	.05	.47
`Intent_to_Quit`	`Intrinsic_motivation`	.24	.07	.24	.07
`Intent_to_Quit`	`Teacher_autonomy`	−.69	<.01	−.57	<.01
`Intent_to_Quit`	`Teacher_competence`	.12	.28	−.02	.83
`Intent_to_Quit`	**`Teacher_relatedness`**	**−.20**	**.18**	**−.28**	**.01**
Predictor factor correlations					
`Teacher_autonomy`	`Teacher_competence`	.77	<.01	.48	<.01
`Teacher_autonomy`	`Teacher_relatedness`	.80	<.01	.52	<.01
`Teacher_relatedness`	`Teacher_competence`	.66	<.01	.42	<.01

Note: Bold values highlight differences in the results of statistical significance tests; the *p*-values are based on the significance tests of unstandardized parameters; *b** are standardized coefficients.

Furthermore, the rotation method and epsilon value can be selected. Researchers can choose 'target rotation' to mimic the confirmatory approach of CFA by specifying the non-target factor loadings to be as close as the researcher's preference (typically zero). However, in this example, we use geomin rotation with an epsilon value of .50 as suggested in the ESEM literature (Marsh et al., 2014; Morin, 2023; Morin et al., 2013), which can be requested in the `sem` function via this code[2]:

```
rotation.args = list(geomin.epsilon = 0.50)
```

All additional code needed for the structural part is principally the same when running CFA/SEM in *lavaan*.

[2]Note that at the time of writing, *lavaan* (0.6-17) does not (yet) support different target specifications for different measurement blocks. For target rotation, only a single measurement block is currently supported. Nevertheless, we have provided the syntax needed to run a simple ESEM model using target rotation in *lavaan* in the same OSF link. It can be seen in the file named 'Target_rotation_example.R', and the associated data is 'Target_rotation_example.dat'.

3.2.3 | Results

To estimate the models, we applied the robust version of maximum likelihood (MLR). To evaluate the quality of the models, we assessed the robust chi-squared statistic, χ^2 (Yuan & Bentler, 2000), its degrees of freedom, and its *p*-value. Other model fit indices were used such as the comparative fit index (CFI), TLI, and RMSEA with its 90% confidence interval. The CFI, TLI, and RMSEA values reported in the two examples are the robust version (Brosseau-Liard et al., 2012; Brosseau-Liard & Savalei, 2014). The results in Table 1 indicate that both the CFA-based structural model and set-ESEM structural model provided acceptable to good levels of goodness of fit. Because our focus in this methodological paper is on examining differences in the structural relationships between CFA-based and set-ESEM structural models, we do not provide a discussion of the measurement model. However, the reader can easily run our models using the shared data and R syntax to obtain full results.

Table 2 shows the path coefficients of the two models, with bold values highlighting significant differences. First, although both models provide acceptable fits to the data, the set-ESEM structural model fits the data better (ΔTLI/CFI = +.01). Furthermore, the Akaike information criterion (AIC), BIC and sample-size-adjusted BIC indices in the CFA-based model were respectively 15,985.43, 16,226.27, and 16,013.84, whereas the ESEM-based model provides the following results: AIC = 15,971.25, BIC = 16,298.37, and sample-size-adjusted BIC = 16,009.84. Lower values of the AIC and sample-size-adjusted BIC in the ESEM-based model suggest a better fit for the ESEM-based model. Moreover, factor correlations of exogenous latent variables are likely to be inflated (i.e., positively biased) in the CFA-based structural model (Shao et al., 2022). For instance, the correlation between `Teacher_autonomy` and `Teacher_relatedness` was .80 in the CFA-based structural model but .51 in the set-ESEM structural model (i.e., $\Delta r = .29$). This, as we will see, can have a determinantal effect on the estimation of model parameters due to possible collinearity resulting from misspecification of the measurement model. For example, some paths increase significantly in the set-ESEM compared to the CFA such that what was non-significant becomes significant. For example, in the CFA-based structural model, the path `Teacher_competence` → `Intrinsic_motivation` is not significant in the CFA-based structural model ($\beta = .08$, $p = .51$) but becomes significant in the set-ESEM structural model ($\beta = .19$, $p = .03$). The same observation applies to the path `Teacher_relatedness` → `Intent_to_Quit` as it is not significant in the CFA-based structural model ($\beta = -.20$, $p = .18$) but becomes significant in the set-ESEM structural model ($\beta = -.28$, $p = .01$). Interestingly, the path `Teacher_competence` → `Intent_to_Quit` is positive in the CFA-based structural model ($\beta = .12$, $p = .28$) and becomes negative in the set-ESEM structural model ($\beta = -.02$, $p = .83$), although both effects are non-significant.

In sum, each model can have different conclusions about the relationship between the latent variables. Given that set-ESEM provides a better fit to the data, this could indicate that the path coefficients can be seen as more meaningful in the set-ESEM structural model than in the CFA-based structural model (due to increased probability of Type II error).

3.3 | Study 2: Using set-ESEM to assess longitudinal models

Study 2 demonstrates the second scenario where the researcher might have constructs collected from different time points. In this empirical example, we measured three constructs at two time points. The first two constructs are autonomy and intrinsic motivation, which we described in Study 1. Both autonomy and intrinsic motivation are drawn from the self-determination theory (Ryan & Deci, 2017). The third construct is self-efficacy, which reflects students' belief about their ability to produce desired results and forestall detrimental ones (Woodrow, 2006). Research in this regard suggests that the three factors (i.e., autonomy, intrinsic motivation, and self-efficacy) work collectively, thus reflecting parallel motives that predict positive outcomes (see Alamer, Al Khateeb, & Jeno, 2023; Alamer & Alrabai, 2023; Alamer, Morin, et al., 2023; Elahi Shrivan & Alamer, 2022; Marsh et al., 2020; Noels, 2023; Ryan & Deci, 2017).

Nevertheless, empirical studies are rare in assessing these factors' long-term effects on L2 achievement. Estimating a model where constructs at time 1 predict their counterpart at time 2 (to control for the stability of a measure over time) while estimating their effect on the outcome variable can be better assessed by set-ESEM than full-ESEM (Marsh et al., 2020). For longitudinal SEM models, it is routinely posited that the uniquenesses of the same item over time need to be covaried (Marsh & Hau, 1996). Additionally, the analyst may want to use measurement invariance to ensure that measurement is stable over time. This can be done by imposing longitudinal equality constraints on factor loadings. This is possible in the long-format code for set-ESEM in *lavaan*. Below is an example of imposing equality constraints for the four items of *self-confidence* at time 1 and time 2. Notice the `a*`, `b*`, `c*`, and `d*` labels, as they express equality constraints on primary loadings. The same variable gets the same label across the two time points. If two parameters have the same label, they will be estimated to be equal. Consistent with the CFA literature on measurement equivalence/invariance, one must still set the latent scale for a 'reference' group or occasion (time 1 in our example), but the remaining ones can be freely estimated. Notice that the missing value, `NA*`, is used to free the factor variance. It is important to note that the only way (in *lavaan*) to preserve the equality constraints across blocks in a rotated solution is to equate all loadings, so that the entire set of loadings in the time 2 block are the same as those in the time 1 block (thus, each block's rotation yields the same solution as the other). Although partial invariance may be a more realistic expectation, partial-invariance constraints would not be preserved in a rotated solution. This method is illustrated using *self-confidence* items as an example as follows:

```
efa("time1")*SelfConfidenceT1 =~ a*SelfConf1T1 + b*SelfConf2T1 +
c*SelfConf3T1 + d*SelfConf4T1 + e*Intr1T1 + f*Intr2T1 + g*Intr3T1 +
h*Auton1T1 + i*Auton2T1 + j*Auton3T1 + k*Auton4T1

efa("time2")*SelfConfidenceT2 =~ a*SelfConf1T2 + b*SelfConf2T2 +
c*SelfConf3T2 + d*SelfConf4T2 + e*Intr1T2 + f*Intr2T2 + g*Intr3T2 +
h*Auton1T2 + i*Auton2T2 + j*Auton3T2 + k*Auton4T2

##freeing factor variance at Time 2
SelfConfidenceT2 ~~ NA*SelfConfidenceT2
```

A benefit of using measurement invariance of set-ESEM could be that it reduces the solution to a single ESEM block which is more parsimonious and often of interest for longitudinal analysis. However, to maintain focus on the standard application of CFA and set-ESEM, we only report the results of the fit indices of the measurement invariance models and do not discuss their path coefficients. Nevertheless, we provide the R syntax on the OSF repository so the reader can reproduce the full results. To test a longitudinal set-ESEM model, we only estimate cross-loadings between the factors at the same time point. For instance, time 1 items for autonomy, intrinsic motivation, and self-efficacy are only cross-loaded on different factors at time 1; time 1 items were constrained to have zero loadings on time 2 constructs. Similarly, time 2 items had cross-loadings on each of the time 2 factors, but not time 1 factors (see Figure 3). Note that uniqueness correlations between the same item at time 1 and time 2 are estimated but not shown in the figure for ease of reading (only one uniqueness correlation example is shown: that between `Aut1` and `Aut1T2`).

3.3.1 | Measures

To assess autonomy, four items from the previously explained BPN-L2 scale were used. An example item was presented in Study 1. To assess intrinsic motivation, three items from the SDT-L2 scale (Alamer, 2022) were adopted. Self-confidence was measured by four items adopted from Dörnyei and Ushioda (2021; see also Dörnyei & Ryan, 2015) as explained in Study 1. The measures are based on a

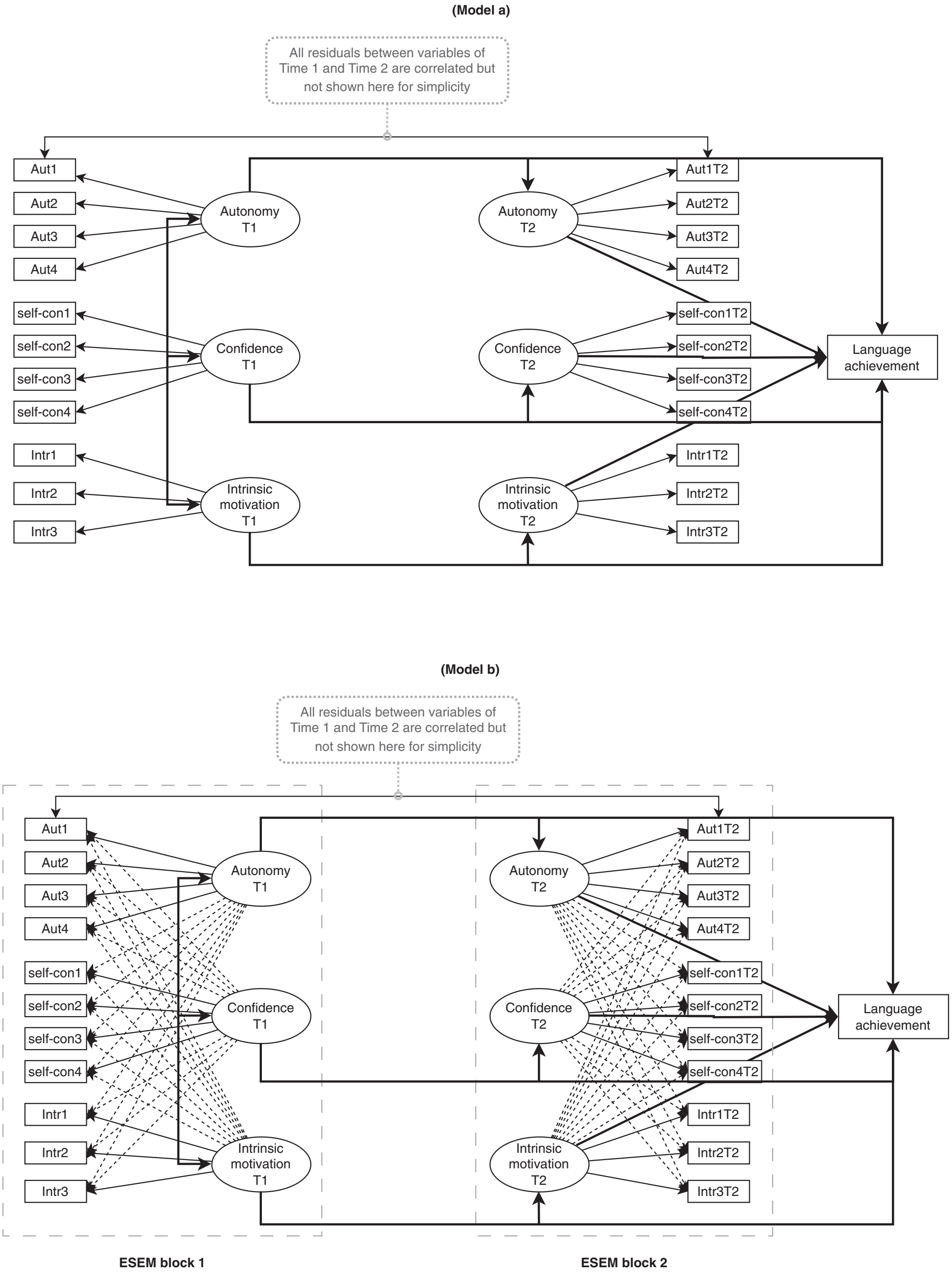

FIGURE 3 A CFA-based longitudinal model (model A) and longitudinal set-ESEM (model B).

five-point Likert response format ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Language achievement was measured by students' placement test of English as a second language. The test includes items on the four skills of vocabulary, grammar, reading, and writing. The test is primarily developed via consulting *Unlock*, an English language course developed by the University of Cambridge (Ostrowska et al., 2021). The test format varies but mainly consists of multiple-choice and fill-in-the-blank items. A simulated sample is included in the OSF repository. The total mark for the test is 20 and the lowest is zero ($M = 14.2$, $SD = 4.7$).

3.3.2 | Results

We estimated the set-ESEM-based and CFA-based structural models using MLR. The results of the two pairs of models are presented in Table 3, suggesting that both provided a satisfactory fit to the data. Note that these results correspond to the metric invariance models due to the fact they fit as well as the configural models. However, the set-ESEM-based model provided a better fit to the data (ΔTLI/CFI = +.01). Further, the AIC, BIC, and sample-size-adjusted BIC indices in the CFA-based model were respectively 23,782.68, 24,034.45, and 23,834.55, while the ESEM-based model provided the following results: AIC = 23,764.25, BIC = 24,067.98 and sample-size-adjusted BIC = 23,826.82. Lower values of the AIC and sample-size-adjusted BIC in the ESEM-based model suggest a better fit for the ESEM-based model.

As can be seen from Table 4, compared to the CFA-based structural model, factor correlations of exogenous latent variables are significantly lower in the set-ESEM-based structural model. For instance, the correlation between `Self_Confidence_T1` and `Autonomy_T1` was .55 in the CFA-based structural model but .29 in the set-ESEM-based structural model (i.e., $\Delta r = .26$). These smaller factor correlation values impacted the prediction (effects) in the structural model, as we will explain next. Specifically, some path coefficients have different effect sizes and p values across the two models. For

TABLE 3 Model fit indices for the CFA-based and set-ESEM longitudinal models.

Model	χ^2	*df*	RMSEA	RMSEA 90% confidence interval	CFI	TLI
CFA-based model	346.240*	205	.044	(.036, .052)	.951	.940
CFA-based model with measurement invariance	353.480*	213	.043	(.035, .050)	.952	.943
Set-ESEM-based model	283.981*	173	.042	(.033, .050)	.962	.945
Set-ESEM-based model with measurement invariance	314.928*	200	.040	(.031, .048)	.961	.950

*Significant at $p < .01$.

TABLE 4 Path coefficients in the two metric-invariance models.

Dependent	Predictor	CFA-based results		Set-ESEM results	
		b^*	p-Value	b^*	p-Value
`Self_Confidence_T2`	`Self_Confidence_T1`	.66	<.01	.66	<.01
`Intrinsic_T2`	`Intrinsic_T1`	.61	<.01	.55	<.01
`Autonomy_T2`	`Autonomy_T1`	.45	<.01	.42	<.01
`L2_achievement`	`Self_Confidence_T1`	.30	<.01	.27	<.01
`L2_achievement`	**`Intrinsic_T1`**	**.10**	**.07**	**.14**	**<.01**
`L2_achievement`	`Autonomy_T1`	.11	.05	.16	<.01
`L2_achievement`	`Self_Confidence_T2`	.39	<.01	.40	<.01
`L2_achievement`	**`Intrinsic_T2`**	**.08**	**.09**	**.12**	**.01**
`L2_achievement`	`Autonomy_T2`	.23	<.01	.26	<.01
Factor correlation					
`Intrinsic_T1`	`Self_Confidence_T1`	.42	<.01	.26	<.01
`Intrinsic_T1`	**`Autonomy_T1`**	**.18**	**.01**	**.06**	**.12**
`Self_Confidence_T1`	`Autonomy_T1`	.55	<.01	.29	<.01

Note: Bold values highlight differences in the results of statistical significance tests; the p-values are based on the significance tests of unstandardized parameters; b^* = standardized coefficients.

instance, the CFA-based structural model suggests that the effect of intrinsic motivation at both time 1 and time 2 is non-significant (e.g., `Intrinsic_T1 → L2_achievement`, $\beta = .10$, $p = .07$, and `Intrinsic_T2 → L2_achievement`, $\beta = .08$, $p = .09$) while in the set-ESEM-based structural model it is significant (`Intrinsic_T1 → L2_achievement`, $\beta = .14$, $p < .01$, and `Intrinsic_T2 → L2_achievement`, $\beta = .12$, $p = .01$). Having path coefficients significant in the set-ESEM-based structural model but not in the CFA-based structural model leads to different conclusions about the prediction (effects) between the variables.

4 | DISCUSSION

In this methodological demonstration, we briefly introduced ESEM in general and set-ESEM in particular. Although set-ESEM is mostly used to evaluate measurement models in the literature, we were particularly interested in elaborating on the utility of set-ESEM in structural models. Specifically, we discussed conceptual justifications for endorsing set-ESEM in structural models over the restrictive CFA model, and even for fully relaxed ESEM specifications. Set-ESEM is a balance between full-ESEM and CFA; it makes it possible to specify mini (independent) sets of ESEM in one solution (Marsh et al., 2020). There are cases where the fully relaxed ESEM might be rather unnecessary, inappropriate, or technically infeasible. We presented two potential applications in which set-ESEM should be selected over full-ESEM. In both applications, we compared the results of set-ESEM with CFA.

Our empirical examples suggest that set-ESEM provided a more meaningful representation of the data than CFA. First, although the CFA models appeared to fit the data without applying exploratory modifications, the factor correlations were systematically higher between the exogenous latent variables in both Study 1 and Study 2. Such high correlations seem to lead to distorted, and thus biased, path coefficients in the structural model due to possible multicollinearity (Mai et al., 2018; Morin, 2023). Our findings illustrate that this concern may be relevant in practice.

With regard to Study 1, path coefficients change substantially across CFA-based and set-ESEM structural models, leading to dissimilar conclusions about the effects between the variables. For example, the statistical significance of the effect *perception about the teacher as providing competence* on students' *intrinsic motivation* differs between the two models. The CFA-based model suggested that this effect is not significant, while set-ESEM suggests a significant effect. Consequently, the researcher would end up supposing that the role of the teacher in enhancing students' intrinsic motivation is not present if the CFA-based model is interpreted in isolation from the set-ESEM model. Nonetheless, only the set-ESEM model confirmed this positive effect. In contrast, the CFA-based structural model suggests that *perceiving teachers as providing relatedness* has an insignificant effect on *students' intention to quit studying the language*. In the set-ESEM structural model, the analysis suggests a negative and significant effect. Thus, only the set-ESEM model provides evidence of the important role teachers can play in helping students maintain their studies and enrolment (Alamer & Al Khateeb, 2023; Alamer, Al Khateeb, & Jeno, 2023; Noels, 2023).

Similarly to Study 1, Study 2 path coefficients varied substantially across CFA-based and set-ESEM structural models, leading to a divergence in conclusions about longitudinal relationships. For example, the CFA-based structural model suggests that *intrinsic motivation* at time 1 and time 2 has little prediction relevance for students' later *L2 achievement*. So, one can conclude a trivial long- and short-term effect of intrinsic motivation on subsequent L2 achievement. However, this conclusion contradicts the mainstream argument of the role of intrinsic motivation in achievement (Alamer, 2022; Alamer & Alrabai, 2023; Horwood et al., 2021).

In short, we found that in the two empirical examples, path coefficients can be non-significant in the CFA-based structural model but significant in the set-ESEM structural model. This is critical to consider when drawing inferences for applied research. However, one might ask which path coefficient should be trusted. We think that this question should be addressed by inspecting the goodness of fit of the models and ensuring an absence of multicollinearity. In both our examples, set-ESEM models fit the data better and the correlations among factors are systematically smaller in support of the factor's distinctiveness, suggesting lower Type II error rates for the path coefficients in the structural model.

4.1 | Methodological remarks

To summarize, below are some of the advantages of using set-ESEM:

- It achieves an optimal balance between parsimony (it is more parsimonious than full-ESEM) and better goodness of fit (it often fits better than CFA).
- It allows distinctive theoretical constructs to be included (but separated) in one model (compared to full-ESEM).
- When target rotation is used, set-ESEM achieves a confirmatory approach (compared to mechanical rotations such as geomin).
- It makes it possible to test structural models that full-ESEM might be unable to test.
- It provides more realistic factor correlations (than CFA), leading to a better assessment of discriminant validity.
- When the measurement model is turned into a structural model, the effects (path coefficients) can be perceived as less attenuated and more accurate, with lower Type II error rates (compared to CFA).

Set-ESEM has paved the way for further applications of SEM models, particularly when structural models (i.e., effects between variables) are involved. However, it has some limitations. For example, although set-ESEM is more parsimonious than full ESEM, it is still less parsimonious than CFA. For that reason, if factor correlations and model fit indices are very similar for CFA and set-ESEM, the CFA model should be preferred based on parsimony. However, when non-ignorable cross-lodaings are suggested by conceptual theory, set-ESEM might be preferred. In addition, the present study has not covered all of the benefits of set-ESEM and the scenarios in which it can be used. More research is warranted to expand on this tutorial paper (also see Marsh et al., 2020; Morin, 2023).

5 | CONCLUSION

In this paper, we have elaborated on the importance of set-ESEM in studying structural models. We also present a short tutorial on how to apply the method using the free software package *lavaan* (Rosseel, 2012) in the R environment. Our recommendation is to adopt set-ESEM in studying both measurement and structural models when the full-ESEM is technically inappropriate or lacks flexibility. We explained this through two scenarios using empirical examples. One can compare the results of set-ESEM to those of CFA to study model fits and model parameter differences. In our case, we found that set-ESEM outperformed the CFA-based structural models in terms of goodness of fit and prior support for path coefficients. Due to increased factor correlations among the exogenous latent variable in the CFA-based structural models, the effects in the structural model appeared to be attenuated. For that reason, some path coefficients were only significant in the set-ESEM models. Although CFA-based models provided an acceptable (and moderately good) fit to the data, they were not free of collinearity issues that led to possible bias in model parameters (Shao et al., 2022; Tabachnick & Fidell, 2023). As such, if we had not conducted set-ESEM but relied solely on CFA-based structural models, we would have come up with different effects between variables, and thus distinct implications from the same data. To this end, we provided two illustrative examples as a showcase for set-ESEM and commented briefly on the results of the competing models. To make this tutorial paper practical, we provide the syntax needed for the R package *lavaan* to facilitate the implementation of set-ESEM. Most importantly, we share the data used in both studies in the open repository OSF so that readers can train and replicate our analysis.

AUTHOR CONTRIBUTIONS

Herb Marsh: Methodology; formal analysis; supervision; writing – review and editing; conceptualization. **Abdullah Alamer:** Conceptualization; writing – original draft; methodology; data curation; investigation; writing – review and editing; formal analysis; project administration; funding acquisition.

ACKNOWLEDGEMENTS

The authors warmly thank the developer of the *lavaan* package, Yves Rosseel, and the associate editor, Terrence Jorgensen, for their help in reading our manuscript and thoroughly checking the R code needed to run the models. Their invaluable feedback and contributions substantially improved the paper. Open access publishing facilitated by University of New South Wales, as part of the Wiley - University of New South Wales agreement via the Council of Australian University Librarians.

FUNDING INFORMATION

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Grant No. 5588).

CONFLICT OF INTEREST STATEMENT

Authors declare that there are no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data, measures, scale items, and R code used for the two studies are included in the OSF repository. None of the studies was preregistered. To access all materials, visit https://osf.io/mhqdx/?view_only=93b57af843254c3e8fbbd31a585123a8.

ORCID

Abdullah Alamer https://orcid.org/0000-0003-4450-0931

REFERENCES

Alamer, A. (2021a). 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. (2021b). Grit and language learning: Construct validation of L2-grit scale and its relation to later vocabulary knowledge. *Educational Psychology*, *41*, 544–562. https://doi.org/10.1080/01443410.2020.1867076

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

Alamer, A., & Al Khateeb, A. (2023). Effects of using the WhatsApp application on language learners motivation: A controlled investigation using structural equation modelling. *Computer Assisted Language Learning*, *36*, 149–175. https://doi.org/10.1080/09588221.2021.1903042

Alamer, A., Al Khateeb, A., & Jeno, L. (2023). Using WhatsApp increases L2 students' self-motivation and achievement, and decreases learning anxiety: A self-determination theory approach. *Journal of Computer Assisted Learning*, *39*, 417–431. https://doi.org/10.1111/jcal.12753

Alamer, A., & Alrabai, F. (2023). The causal relationship between learner motivation and language achievement: New dynamic perspective. *Applied Linguistics*, *44*, 148–168. https://doi.org/10.1093/applin/amac035

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*, *44*, 1477–1500. https://doi.org/10.1017/S0272263121000863

Alamer, A., Morin, J. S. A., Alrabai, F., & Alharfi, A. (2023). Introducing the basic psychological needs frustration in second language scale (BPNF-L2): Examining its factor structure and effect on L2 motivation and achievement. *Acta Psychologica*, *240*, 104021. https://doi.org/10.1016/j.actpsy.2023.104021

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 397–438. https://doi.org/10.1080/10705510903008204

Brosseau-Liard, P. E., & Savalei, V. (2014). Adjusting incremental fit indices for nonnormality. *Multivariate Behavioral Research*, *49*(5), 460–470. https://doi.org/10.1080/00273171.2014.933697

Brosseau-Liard, P. E., Savalei, V., & Li, L. (2012). An investigation of the sample performance of two nonnormality corrections for RMSEA. *Multivariate Behavioral Research*, *47*(6), 904–930. https://doi.org/10.1080/00273171.2012.715252

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. Routledge.

Dörnyei, Z., & Ushioda, E. (2021). *Teaching and researching motivation*. Routledge.
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. https://doi.org/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*, 1–17. https://doi.org/10.1080/01434632.2022.2075002
Faraci, P., Bottaro, R., Nasonte, G., Valenti, G. D., & Craparo, G. (2023). A new short version of Internet Gaming Disorder-20 (IGD-20): An exploratory structural equation modeling. *Clinical Neuropsychiatry*, *20*(3), 205–218.
Gegenfurtner, A. (2022). Bifactor exploratory structural equation modeling: A meta-analytic review of model fit. *Frontiers in Psychology*, *13*, 1037111. https://doi.org/10.3389/fpsyg.2022.1037111
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. *Journal of Experimental Education*, *83*, 51–82. https://doi.org/10.1080/00220973.2013.876231
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. https://doi.org/10.1037/edu0000664
Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. https://doi.org/10.1080/10705519909540118
Kruk, M., Pawlak, M., Shirvan, M. E., & Soleimanzadeh, S. (2023). Revisiting boredom in practical English language classes via exploratory structural equation modeling. *Research Methods in Applied Linguistics*, *2*, 100038. https://doi.org/10.1016/j.rmal.2022.100038
Liu, E., Wang, J., & Bai, S. (2022). Validation of L2 grit among Chinese EFL high school students and its enduring effect on achievements: A bifactor model approach. *Frontiers in Psychology*, *13*, 971495. https://doi.org/10.3389/fpsyg.2022.971495
Lounsbury, J. W., Saudargas, R. A., & Gibson, L. W. (2004). An investigation of personality traits in relation to intention to withdraw from college. *Journal of College Student Development*, *45*(5), 517–534.
Mai, Y., Zhang, Z., & Wen, Z. (2018). Comparing Exploratory Structural Equation Modeling and Existing Approaches for Multiple Regression with Latent Variables. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(5), 737–749. https://doi.org/10.1080/10705511.2018.1444993
Marsh, H., 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., & Hau, K. (1996). Assessing goodness of fit: Is parsimony always desirable? *Journal of Experimental Education*, *64*, 364–390. https://doi.org/10.1080/00220973.1996.10806604
Marsh, H., Morin, A. J., Parker, P., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*, 85–110.
Marsh, H., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 439–476. https://doi.org/10.1080/10705510903008220
Morin, A. J. (2023). Exploratory structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (2nd ed.). Guilford.
Morin, A. J., Marsh, H., & Nagengast, B. (2013). Exploratory structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 395–436). Information Age Publishing.
Noels, K. (2023). Self-determination theory and language learning. In *The Oxford handbook of self-determination theory* (pp. 619–637). Oxford University Press.
Ostrowska, S., O'Neill, R., Westbrook, C., Sowton, C., White, N. M., Dimond-Bayir, S., Lansford, L., Pathare, E., & Pathare, G. (2021). *Unlock*. Cambridge University Press. https://www.cambridge.org/gb/cambridgeenglish/catalog/skills/unlock
Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36. https://doi.org/10.18637/jss.v048.i02
Ryan, R., & Deci, E. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publications.
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*, 901412. https://doi.org/10.3389/fpsyg.2022.901412
Spearman, C. (1904). General intelligence' objectively determined and measured. American. *Journal of Psychology*, *5*, 201–293.
Swami, V., Maïano, C., & Morin, A. J. (2023). A guide to exploratory structural equation modeling (ESEM) and bifactor-ESEM in body image research. *Body Image*, *47*, 101641. https://doi.org/10.1016/j.bodyim.2023.101641
Tabachnick, B. G., & Fidell, L. S. (2023). *Using multivariate statistics* (7th ed.). Pearson Education.
Thurstone, L. L. (1935). *The vectors of the mind*. University of Chicago Press.
Thurstone, L. L. (1947). *Multiple factor analysis*. University of Chicago Press.
Vallerand, R. (2015). *The psychology of passion: A dualistic model*. Oxford University Press.
Van Zyl, L. E., & ten Klooster, P. M. (2022). Exploratory structural equation modeling: Practical guidelines and tutorial with a convenient online tool for Mplus. *Frontiers in Psychiatry*, *12*, 2397. https://doi.org/10.3389/fpsyt.2021.795672
Woodrow, L. (2006). A model of adaptive language learning. *Modern Language Journal*, *90*, 297–319.

Yuan, K.-H., & Bentler, P. M. (2000). 5. Three likelihood-based methods for mean and covariance structure analysis with non-normal missing data. *Sociological Methodology*, *30*(1), 165–200. https://doi.org/10.1111/0081-1750.00078

How to cite this article: Marsh, H., & Alamer, A. (2024). When and how to use set-exploratory structural equation modelling to test structural models: A tutorial using the R package lavaan. *British Journal of Mathematical and Statistical Psychology*, *00*, 1–18. https://doi.org/10.1111/bmsp.12336