EAM2025
XI Conference
23RD - 25TH JULY 2025
European Association of Methodology
Spain Tenerife Canary Islands

# Methodological differences between formative-measured and composite variables: a case study using mixed SEM techniques

Cynthia Mª Delgado García, Daniel Ondé Pérez & Jesús Mª Alvarado Izquierdo

UNIVERSIDAD COMPLUTENSE DE MADRID

## Introduction

The difference between causal-formative and composite variables is often neglected in structural equation modelling (Henseler, 2017), difficulting the progress and discussions that focus on causal-formative variables (Bollen & Diamantopoulos, 2017). One problem that both causal-formative and composite variables face is their specification in a model as endogenous variables (Rigdon, 2014; Yu et al., 2022). Temme et al. (2014) proposed a guideline to correctly specify endogenous causal-formative variables under both Covariance Structure Analysis (CSA) and partial least squares (PLS) perspectives. These autors recommend the use of CSA analysis instead of PLS and avoiding the use of aggregated variables (i.e. composites). However, the development of Confirmatory Composite Analysis (CCA, Schuberth et al, 2018) the synthesis theory (Henseler & Schuberth, 2021) strongly supports the use of composites in theory-building research. In addition, CCA differs from the traditional estimation and validation processes of PLS-estimated models (Schuberth, 2020), making it similar to Confimatory Factor Analysis (CFA). This approach can be combined with Temme et al. (2014) using the Henseler-Ogasawara specification, that permits the use of composite variables in endogenous positions and its estimation using CSA (Schuberth, 2023). Also, a refined version of the Henseler-Ogasawara specification exists, improving the convergence issues of the original specification (Yu et al., 2022).

Ondé et al. (2023) proposed a model that studied the relationships between three latent variables: emotional intelligence, meta-comprehension knowledge and oral communication skills. Their chosen model contains causal-formative variables in endogenous positions (although these variables are equivalent to composite variables in PLS, the estimation method used in the paper). Reflecting on the presented advances of causal-formative and composite variables, the model of Ondé et al. (2023) was revaluated.

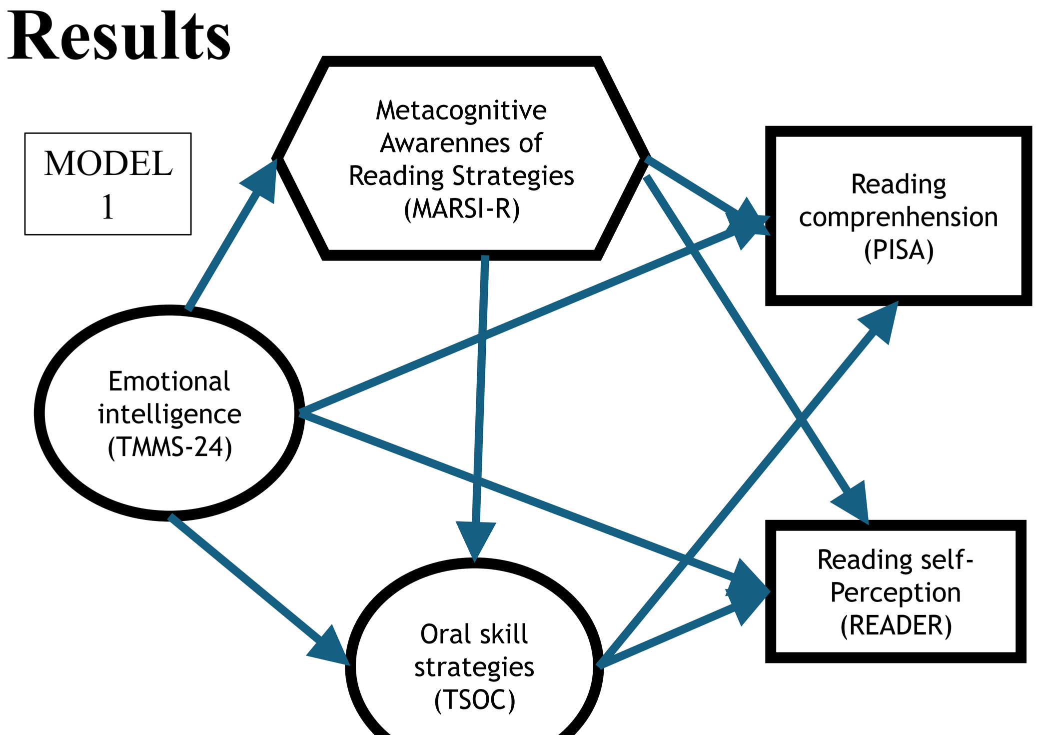
## Objectives

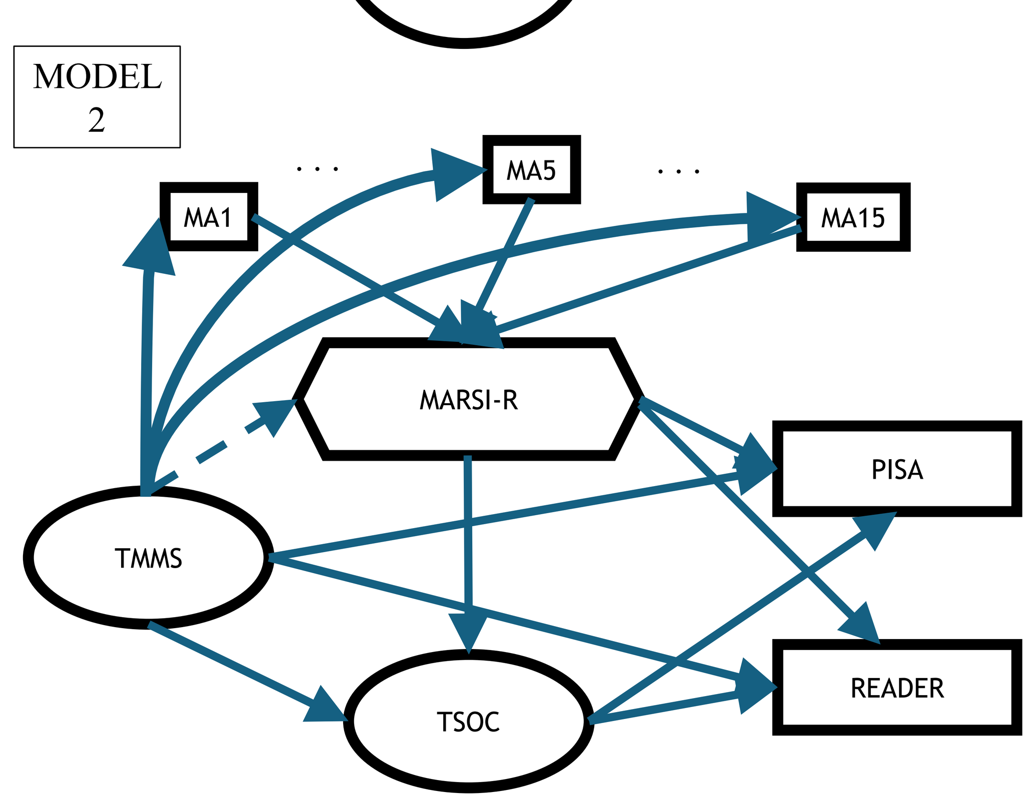
To evaluate the original model and alternative models. In addition, to show some practical differences of modelling causal-formative and composite variables.
Model 1: original
Model 2: TMMS -> MARSI via MARSI items
Model 3: MARSI is a composite (multidimensional)

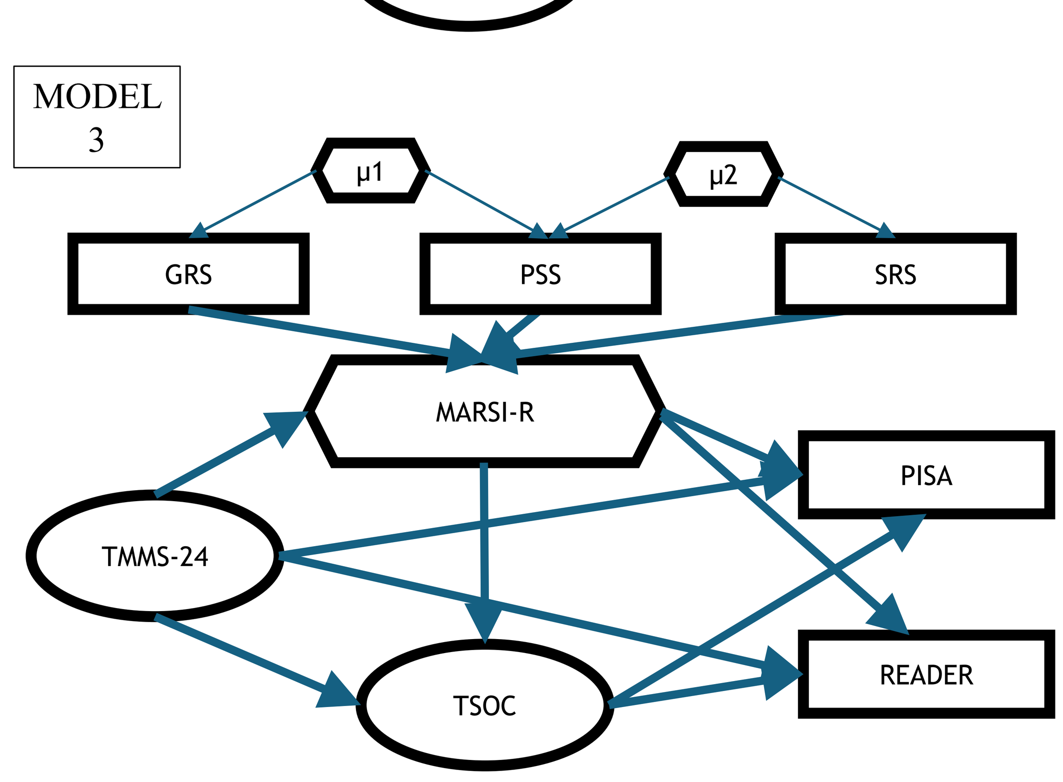
## Methodology

Participants: 327 students (46.2% male and 26.3% female) aged 12–17 years (mean age = 14.5, SD = 1.2).
The original model has one reflective variable (TMMS-24), two causal-formative variables (MARSI-R and TSOC) and two outcomes (PISA and READER items). Here, the "original" model has only one causal-formative variable (MARSI-R) and an extra path from MARSI-R to TSOC was added. Also, the latent variables were regarded as unidimensional.
The packages *seminr* and *lavaan* in R (R Core Team, 2024) were used to estimate the models using CSA (estimator = MLR) and PLS for the first and second model and CSA (estimator = MLR) for the third model.

## Results




The models' paths are similar between type of analysis. However, the difference between CSA and PLS remains the same for Model 2 than for Model 1.

| Model Paths | Model 1: CSA | Model 1: PLS | Model 2: CSA | Model 2: PLS | Model 3: CSA |
|---|---|---|---|---|---|
| TMMS >MARSI | -.75 (p=.85) | .38 (.31-.49) | 0 (2nd step) | 0 (2nd step) | -.39 (p<.00) |
| TMMS >TSOC | .39 (p=.90) | .33 (.28-.41) | .52 (p=.90) | .46 (.37-.55) | .34 (p=.06) |
| MARSI >TSOC | -.79 (p=.20) | .46 (.38-.56) | -.96 (p=.13) | .29 (.19-.38) | 1.25 (p<.00) |
| TMMS >PISA | .02 (p=.95) | .00 (-.13-.12) | .02 (p=.65) | .01 (-.11-.14) | .02 (p=.57) |
| TMMS >READER | .23 (p=.96) | -.05 (-.17-.11) | -.27 (p=.72) | -.05 (-.17-.06) | -16.23 (p=.37) |
| MARSI >PISA | -.07 (p=.24) | .14 (-.02-.30) | -.07 (p=.24) | .03 (-.09-.15) | .15 (p=.03) |
| MARSI >READER | 1 | -.37 (-.56-(-.26) | 1 | -.16 (-.28-(-.03) | -1.64 (p=.13) |
| TSOC >PISA | -.03 (p=.83) | .12 (-.02-.25) | -.01 (p=.91) | .19 (.06-.31) | -.05 (p=.29) |
| TSOC >READER | .67 (p=.51) | -.05 (-.17-.11) | .44 (p=.55) | -.18 (-.31-(-.05) | .68 (p=.25) |

Although the Model 2 is more complex than Model 1, their *df* are 1762 and 1761, respectively. The fit is bad, but it is interesting that the fit seems stable across different specifications of the MARSI variable.

| Fit Indices | Model 1. CSA | Model 2. CSA | Model 3. CSA |
|---|---|---|---|
| Chi-square | 4034.342 (p<.00) | 4036.635 (p<.00) | 4309.883 (p<.00) |
| Robust CFI | .716 | .716 | .694 |
| Robust TLI | .685 | .685 | .682 |
| Robust RMSEA | .066 | .066 | .066 |
| Robust SRMR | .082 | .082 | .085 |

## Discussion

This poster shows alternative ways to specify a model with a non-reflective latent variable in a endogenous position and it tried to show the importance of deciding if any latent variable is causal-formative or a composite. Also, by using CSA and PLS analysis, this poster helps to distinguish the composite model (i.e. variables are artifacts) from composite-based SEM (such as PLS, which can be used to model latent variables).
Following the guideline of Temme at al. (2014), Model 2 should avoid biased parameters estimates and should have correct total effects. In this example, there are no present differences between Model 1 and Model 2. Although in Model 2 the relations between TMMS-24 and MARSI-R items are significant (in both CSA and PLS analysis), the relation of those items with MARSI-R are non-significant (i.e. the total effect from TMMS-24 to MARSI-R is non-significant for both models).
On the other hand, Model 3 has a significant effect from TMMS-24 to MARSI-R and a positive path from MARSI-R to TSOC, whereas Model 1 and 2 have a negative value in this path (using CSA). Taking this into account, Model 3 would seem preferable in a theorical way.
The work presented in this poster have several limitations. First, in CSA, the models suffers from covergence issues (with n=327), so the values are questionable. In second place, these variables are multidimensional (TMMS-24 and MARSI-R have 3 subdimensions and TSOC has five) but they were specified as unidimensional in Models 1 and 2. Model 3 being specified as multidimensional could explain its better results. Lastly, the differences between these ways of specify models should be explored with models that show a good fit.

## References

Bollen, K. A. & Diamantopoulos, A. (2017). In defense of causal-formative indicators: A minority report. *Psychological methods, 22(3)*, 581–596. https://doi.org/10.1037/met000005
Henseler, J. (2017). Bridging Design and Behavioral Research With Variance-Based Structural Equation Modeling. *Journal of Advertising, 46(1)*, 178-192. DOI: 10.1080/00913367.2017.1281780
Henseler, J. & Schuberth, F. (2021). Auxiliary theories. In J. Henseler (ed.), *Composite-based structural equation modeling: Analyzing latent and emergent variables* (pp. 25–37). The Guilford Press, New York.
Ondé, D., Cabellos, B., Gràcia, M., Jiménez, V. & Alvarado, J. M. (2023). The Role of Emotional Intelligence, Meta-Comprehension Knowledge and Oral Communication on Reading Self-Concept and Reading Comprehension. *Education Sciences, 13(12)*, 1249. https://doi.org/10.3390/educsci13121249
Rigdon, E. E. (2014). "Comment on "Improper use of endogenous formative variables". *Journal of Business Research, 67(1),* 2800-2802.
Schuberth, F. (2020). Confrmatory composite analysis using partial least
squares: setting the record straight. *Review of Managerial Science, 15*, 1311-1345. https://doi.org/10.1007/s11846-020-00405-0
Schuberth, F. (2023). The Henseler-Ogasawara specification of composites in structural equation modeling: A tutorial. *Psychological Methods, 28(4)*, 843–859. https://doi.org/10.1037/met0000432
Schuberth, F., Henseler, J. & Dijkstra, T. K. (2018). Confirmatory Composite Analysis. *Frontiers of Psychology (METHODS), 9*. DOI: 10.3389/fpsyg.2018.02541
Temme, D., Diamantopoulos, A. & Pfegfeidel, V. (2014). Specifying formatively-measured constructs in endogenous positions in structural equation models: Caveats and guidelines for researchers. *International Journal of Research in Marketing, 31(3)*, 309-316.
Yu, X., Schuberth, F. & Henseler, J. (2023). Specifying composites in structural equation modeling: A refinement of the Henseler–Ogasawara specification. *Statistical Analysis and Data Mining, 16*, 348–357. https://doi.org/10.1002/sam.11608

FUNDING STATEMENT. This work has been funded by the proyect PID2022-136905OB-C22 / AEI/10.13039/501100011033/ FEDER, UE. Predoctoral fellowship from the Spanish Ministry of Universities (FPU23/02914), granted to Cynthia Mª Delgado García.