Educational Psychology Review (2026) 38:80
https://doi.org/10.1007/s10648-026-10179-8

COMMENTARY

On Model Interpretability and Time-Reversal Testing in Analyses of Reciprocal Effects: Clarifying Modeling Issues in Sorjonen et al.'s (2025a, 2025b) Commentaries

Fernando Núñez-Regueiro[1] · Herbert W. Marsh[2,3] · Reinhard Pekrun[2,4] · Oliver Lüdtke[5,6] · Jiesi Guo[2]

Received: 27 February 2026 / Accepted: 18 May 2026

Abstract

Núñez-Regueiro et al. (2025) recently demonstrated that the alternative models used by Sorjonen et al. (2025a) to reject the reciprocal effects model (REM) contained misspecifications that, when addressed, yielded results consistent with the REM. In a reply to our critique, Sorjonen et al. (2025b) defended their original models as valid modeling approaches. They further proposed time-reversal tests, which they interpreted as contradicting the REM. In this new comment, we adopt a more pedagogical approach to resolve these misunderstandings. First, we explain in details why the models in Sorjonen et al. (2025a) are difficult to reconcile with several principles of structural equation modeling, by yielding improper solutions, impossible or confounded model parameters, untested temporal structures, and omitted adjustments that limit the interpretability of reciprocal effects. Second, we further explain why the renewed claims of validity in Sorjonen et al. (2025b) extend these issues by mobilizing simulation studies that do not align with the models at stake. Third, we demonstrate that their time-reversal testing repeated some of these issues and that a rectified version yields results consistent with predictions based on the REM (https://osf.io/ry6p3/overview?view_only=2af2bc8cd4694433b9024eb2577451b6). Our objective is to advance the debate on reciprocal effects by clarifying the modeling issues associated with Sorjonen et al.'s commentaries and providing additional empirical evidence relevant to evaluating the REM.

Keywords Reciprocal effects model · Model admissibility · Scientific debate · Structural equation modeling · Causal inference

Extended author information available on the last page of the article

Published online: 26 May 2026

Introduction

The recent exchange in this journal surrounding the reciprocal effects model (REM) has been both valuable and constructive. Sorjonen et al. (2025a) initial comment challenged the findings reported in Marsh et al. (2024) by proposing alternative longitudinal models intended to reexamine the reciprocal effects between academic self-concept (ASC) and achievement (ACH). In our first response (Núñez-Regueiro et al., 2025). We sought to build on this contribution by integrating the models proposed by Sorjonen et al. within a broader comparative framework. In doing so, we addressed misspecifications in Sorjonen et al.'s models by applying validated methodologies that were misapplied or implemented in ways inconsistent with established practice in their commentary (Ghisletta & McArdle, 2012; Kenny & Zautra, 1995; Kievit et al., 2018; Muthén & Asparouhov, 2024; Usami et al., 2019). Our comparative framework gave even stronger support for the findings reported in Marsh et al. (2024). We demonstrated this with reproducible results and detailed explanations of each methodology (https://osf.io/84bzp/). These results and methodology were further replicated in a French sample, providing additional support for the REM (Núñez-Regueiro et al., in press). Therefore, the central issue in the debate is not that alternative models of reciprocal effects contradict findings in Marsh et al. (they do not), but that valid specification of these models is central to their interpretation. Nonetheless, in that first response, we placed greater emphasis on a constructive synthesis than on detailed evaluation of modeling choices in Sorjonen et al.'s comment (2025a), in order to advance the scientific debate on the nature of reciprocal effects between ASC and ACH.

In their new reply, however, Sorjonen et al. (2025b) present their earlier modeling approaches as defensible alternatives. As we detail next, their reply rests in part on placing greater emphasis on minor modeling issues we had documented and by using simulation studies to defend them as valid modeling approaches. At the same time, other major misspecifications we had documented, which affect the interpretability of the models, remain unaddressed (Núñez-Regueiro et al., 2025). Furthermore, they propose time-reversal tests of reciprocal effects that appear to contradict a causal interpretation of the REM; yet, these new tests repeat some of the earlier modeling issues. Taken together, these points suggest that our first response (Núñez-Regueiro et al., 2025) did not explain in sufficient detail the modeling issues in Sorjonen et al.'s initial comment (2025a), which may have led Sorjonen et al. (2025b) to consider their misspecified models as defensible and extend them to time-reversal tests.

The main goal of this response is therefore to provide a more detailed and pedagogical clarification of the modeling issues in Sorjonen et al.'s comments (Sorjonen et al. 2025a, b). We believe that clarifying these issues is important to ensure that the debate on the REM remains grounded in more valid and interpretable modeling practices. In this perspective, we first provide in-depth explanations of the modeling issues in Sorjonen et al.'s first comment (2025a), before explaining how they extend to their second comment (Sorjonen et al. 2025b). An overview of major issues is presented in Table 1. A third section provides new evidence that the REM is also robust to time-reversal tests, providing a pattern of results that supports the existence of reciprocal effects between ASC and ACH.

Table 1 Overview of Issues in Models by Sorjonen et al. (2025a, b)

Model	Major Misspecification	Other Misspecifications/Misinterpretations
LCSM (Sorjonen et al. 2025a)	Trait factors (capturing average change scores) were defined starting on wave 1, assuming an impossible change score (ΔX_1)from a non-existent prior wave 0 (i.e., $\Delta X_1 = X_1 - X_0$). This assumption contradicted LCSM methodology (Ghisletta & McArdle, 2012; Kievit et al., 2018).	- Autoregressive (AR) effects were replaced by covariances, leaving within-construct effects unadjusted and compromising the identification of between-construct effects. This contradicts the principles for interpreting LCSM parameters (Ghisletta & McArdle, 2012; Kievit et al., 2018). - The LCSM showed an improper solution (non-positive definite matrix) but was deemed interpretable nonetheless, disregarding model validity assumptions (Brown, 2015; Kline, 2023; Wothke, 1993).
STARTS (Sorjonen et al. 2025a)	Contemporaneous (lag0) effects were allowed to start on wave 1, preventing adjustment for prior states and undermining the estimation of within-person effects. This contradicts principles for estimating lag0 effects (Muthén & Asparouhov, 2024)	The STARTS specified an uneven structure of effects: ACH was given both lag0 and lag1 effects, whereas ASC had only lag1 effects, thereby distorting parameter balance and interpretability. This untested structure contradicted the methodology for contrasting lag0 and lag1 effects (Muthén & Asparouhov, 2024)
ESDM (Sorjonen et al. 2025a)	The ESDM assumed a unidirectional effect from trait ACH to trait ASC ($ACH_{Trait} \rightarrow ASC_{Trait}$), despite both being time-invariant random intercepts. This contradicted principles of causal inference (Bailey et al., 2024; Usami et al., 2019).	The same model fit and parameters are obtained if the path is reversed ($ASC_{Trait} \rightarrow ACH_{Trait}$) or replaced by a covariance; the reported unidirectional effect therefore lacks statistical basis or empirical testing.
LCSM simulation study (Sorjonen et al., 2023)	Claimed to show that autoregressive effects in LCSM create spurious associations, but the simulated data had no temporal structure and, thus, did not model change processes at stake in LCSM methodology.	This simulation study was used to justify their unconventional LCSM (Sorjonen et al. 2025a). However, its conclusions about LCSM bias were invalid because the simulation data were incompatible with the change processes.
"Time-reversed" REM (Sorjonen et al. 2025b)	Construed as a time-reversed version of the REM in Marsh et al. (2024), but actually changes it drastically: the lag1 effect of ASC on ACH is replaced by a lag0 effect; it omits the lag0 effect of ACH on ASC; it omits the autoregressive effects of ASC, etc.	- Reversed time only for ACH (not ASC) - Started contemporaneous effects on wave 1, thereby repeating the same misspecification as in the STARTS model (i.e., lack of adjustment for prior states).

All listed specification issues contradict standard principles of structural equation modeling (i.e., model identification, temporal ordering, and parameter interpretability), raising serious concerns about the interpretability of the resulting estimates. Addressing these issues yields proper solutions (positive-definite matrices), interpretable parameters, and valid tests, all of which corroborate the reciprocal effects originally reported in Marsh et al. (2024)

Why Alternative Models In Sorjonen et al. (2025a) Are Not Interpretable as Specified

Sorjonen et al. (2025a) initially claimed that prior findings reported by Marsh et al. (2024) were not robust to alternative models of reciprocal effects. These alternative models included the latent change score model (LCSM), the stable trait, autoregressive trait, and state model (STARTS), and the extended skill development model

(ESDM). In our first response (Núñez-Regueiro et al., 2025), we argued that the models actually implemented in Sorjonen et al. (2025a) departed from and contradicted the original methodology established for these models (Ghisletta & McArdle, 2012; Kievit et al., 2018; Kline, 2023; Muthén & Asparouhov, 2024; Usami et al., 2019). However, Sorjonen et al.'s reply (2025b) did not fully engage with these issues, perhaps because our earlier response focused primarily on corrected versions of these models. To clarify the debate, it is important to restate these issues explicitly. In this section, we summarize the most critical ones (Table 1) (for other issues, see Núñez-Regueiro et al., 2025).

LCSM The LCSM implemented by Sorjonen et al. (2025a) raised several specification concerns, the most critical being that the trait factors (i.e., average change scores in ASC and ACH) were specified to begin in wave 1. In a standard LCSM, change scores (Δ) represent differences between successive waves ($\Delta X_2 = X_2 - X_1$, $\Delta X_3 = X_3 - X_2$, etc.), and these change scores define the trait factors (Ghisletta & McArdle, 2012; Kievit et al., 2018). Since wave 0 does not exist, change scores cannot exist in wave 1 (i.e., ΔX_1 is not identifiable). This means that the trait factors can only be defined starting on wave 2, when the change scores actually begin (i.e., ΔX_2 and onwards). By contrast, when the trait is allowed to load on wave 1—as in Sorjonen et al.'s LCSM, the model is forced to treat the score on wave 1 (X_1) as a change score (i.e., ΔX_1), which is itself not defined. This specification distorts the structure of variance-covariance in the data. This is one reason why the LCSM in Sorjonen et al. yielded non-positive definite variance-covariance matrices: The estimation algorithm had to allocate "variance" to a change score factor (i.e., ΔX_1) that did not exist in the data. We demonstrated (Núñez-Regueiro et al., 2025) that once these and other issues are corrected, the LCSM yields a proper solution (positive definitive variance-covariance matrix) and shows positive effects between ASC and ACH, consistent with the findings reported by Marsh et al. (2024).

STARTS The STARTS model implemented by Sorjonen et al. (2025a) also raised specification concerns, notably regarding the inclusion of contemporaneous effects. Contemporaneous effects represent predictive relations occurring within the same wave (lag0 effect), for example, when ASC on wave 2 predicts ACH on wave 2 (effect $ACH_2 \rightarrow ASC_2$). Identifying contemporaneous effects requires adjusting for the influence of prior states, otherwise the residual variance from earlier occasions (e.g., effect $ASC_1 \rightarrow ASC_2$) becomes confounded with the contemporaneous effect estimated at time t (e.g., effect $ACH_2 \rightarrow ASC_2$). In practice, this means that the contemporaneous effect can only be estimated beginning on wave 2 (Muthén & Asparouhov, 2024). In contrast, Sorjonen et al. specified the contemporaneous effects to begin on wave 1. As a result, their STARTS model effectively confounded within-constructs dynamics (effect $ASC_T \rightarrow ASC_{T+1}$) with contemporaneous effects (effect $ACH_{T+1} \rightarrow ASC_{T+1}$), which biases parameter estimates and undermines the interpretation of results. When the contemporaneous effects are introduced beginning on wave 2 (Núñez-Regueiro et al., 2025), the STARTS model shows positive contemporaneous effects of ACH on ASC, and positive lagged effects of ASC on ACH, consistent with the findings reported by Marsh et al. (2024).

ESDM The ESDM was introduced by Sorjonen et al. (2025a) as a new theoretical framework to explain directional ordering in the reciprocal effects model (REM) and in other contexts (Sorjonen et al. 2025c). In the ESDM, each construct (ASC and ACH) is represented as a stable, trait-like factor via random intercepts. Because these trait factors do not change over time (they are intercepts), one can model their association, but not their effects on each other. This is because causal inference in longitudinal modeling requires temporal or experimental variation to establish directionality of effects (Bailey et al., 2024; Usami et al., 2019). Yet, despite lacking such variation, the ESDM imposes a unidirectional effect where trait ACH predicts trait ASC ($ACH_{Trait} \rightarrow ASC_{Trait}$), interpreted as a unidirectional effect. From a causal modeling perspective (Bailey et al., 2024; Kline, 2023; Usami et a., 2019), such interpretation is problematic because identical fit and parameter estimates are obtained if the effect is reversed ($ASC_{Trait} \rightarrow ACH_{Trait}$) or replaced by a correlation ($ACH_{Trait} \leftrightarrow ASC_{Trait}$), as we showed before (Núñez-Regueiro et al., 2025). In other words, the direction of this path is not empirically identified. The ESDM thus exemplifies a situation in which a model can be specified, yet its central claim (that trait ACH predicts trait ASC) cannot be tested empirically. We argued in favor of a more agnostic form of the ESDM replacing its directional path by a covariance between traits, namely the random-intercept third-variable confounder (or time-varying covariate; RI-LTVC; Kenny & McCoach, 2025; Núñez-Regueiro et al., 2025). This model can be used to examine whether covariations between state-level ACH and ASC could be explained by an unobserved confounder. This RI-LTVC had the second worst fit to the data compared to other models and a 0% probability of reflecting the true relations between ASC and ACH (Núñez-Regueiro et al., 2025), providing no support for a time-varying confounder explanation and indicating that the findings reported by Marsh et al. (2024) remain robust under this framework as well.

Clarifying Key Modeling Arguments in Sorjonen et al. (2025b)

The reply by Sorjonen et al. (2025b) overlooks the major modeling issues documented before (Table 1) and instead focuses on other, more minor points we had uncovered (Núñez-Regueiro et al., 2025). Although these issues are less central to the models' validity, they warrant clarification, as they are used to justify the proposed modeling approaches. In this section, we address these additional points, focusing on their implications for the debate on the REM and for the analysis of reciprocal effects between ASC and ACH.

On the Distinction Between Modeling Choices and Specification Errors

Sorjonen et al. (2025b) argue that the debate reflects differences in modeling choices and that their previous models can be viewed as equally valid analytical perspectives as those we proposed in our response (Núñez-Regueiro et al., 2025). This argument is important to the debate and requires clarification. Specifically, the issues identified in our previous response (Núñez-Regueiro et al., 2025) and, detailed in Table 1, depart from SEM logic and compromise parameter interpretability. For example, allowing contemporaneous effects to begin in wave 1 of the STARTS is not a modeling choice but a specification error that compromises the model as a whole. This distinction between modeling choice and specification error may not have been fully explicit in our earlier response but is now clarified in the preceding section.

On the Specification and Interpretation of the LCSM

In their reply, Sorjonen et al. (2025b) defend their LCSM specification on two main grounds. First, they argue that the non-positive-definite matrix of their LCSM does not preclude the interpretation of its parameter estimates. However, this position is difficult to reconcile with standard SEM principles, according to which improper solutions indicate model misspecification and invalidate the interpretation of parameters (Brown, 2015; Kline, 2023; Wothke, 1993). The presence of a non-positive-definite matrix may instead reflect the specification issues documented previously (e.g., inclusion of an unidentified change-score and removal of AR effects; Table 1).

Second, they justify replacing autoregressive (AR) effects between scores (i.e., how past scores of a construct predict its future scores) by covariances (i.e., non-predictive relations) by referring to the results of an earlier simulation study (Sorjonen et al., 2023). In that study, the authors argued that including AR effects in an LCSM (i.e., $X_1 \rightarrow X_2$) could create illusory effects of scores on change scores (i.e., $X_1 \rightarrow \Delta X_2$), which could be avoided by specifying covariance instead. However, inspection of their simulation design reveals that their simulated data did not reflect change scores (nor an LCSM), because it lacked a temporal structure: Both X_1 and X_2 were independently generated from the same underlying latent variable X_{true} (i.e., $X_1 = X_{true} + e_1$ and $X_2 = X_{true} + e_2$, where e_1 and e_2 are random noise), not from a latent change process involving accumulating trait effects or cross-constructs effects (see OSF repository: https://osf.io/pw8ym/). Consequently, these simulations do not demonstrate that AR effects can distort LCSM parameters; rather, they show that dynamic parameters in an LCSM are uninterpretable when applied to non-temporal data. These simulations are therefore not appropriate for justifying the replacement of AR effects with covariances in the LCSM. Established LCSM methodologies (Ghisletta & McArdle, 2012; Kievit et al., 2018; Usami et al., 2015) concur instead that without AR effects, the residual variance arising from past states of one construct (e.g., effect of X_2 on X_3) remains unadjusted and becomes confounded with the cross-lagged effects from the other construct (e.g., of Y_2 on X_3), leading to biased parameters in both sign and magnitude.

On the Limits of the SIM-NEW Simulation as Evidence Against the REM

Finally, Sorjonen et al. (2025b) claim that their newly generated dataset (SIM-NEW) shows that effects such as those in the REM can appear even without causal relations between ASC and ACH. However, this simulation relies on a set of assumptions that depart in important ways from the original context in Marsh et al. (2024). First, their SIM-NEW data are entirely simulated, sharing with the PALMA dataset of Marsh et al. (2024) only the number of participants, despite being used for comparison with it. Second, the simulation model assumes that both ASC and ACH are fully determined by a single latent "ability" factor, thereby collapsing two distinct psychological constructs into one, despite meta-analytic evidence showing that they are only moderately correlated (around r =.30) (Freund & Kasten, 2012; Zell & Krizan, 2014). Third, the analysis of the SIM-NEW data relies on a time-reversed model (Fig. 1, A), discussed in the next section, which does not correspond to the REM tested in Marsh et al. (2024) and therefore does not constitute a comparison test. Taken together, these assumptions limit the relevance of both the simulated data and their analysis for informing the debate on reciprocal effects between ASC and ACH, as the conclusions reflect primarily idiosyncrasies of the data-generating model and of the time-reversal testing, rather than properties of the REM itself.

The Time-Reversal Testing Further Corroborates the REM

Time-reversal tests were introduced for neuroimaging analysis (Haufe et al., 2013) to assess the Granger-causal strength of prospective effects identified in noisy contexts. The rationale of this technique is that, since causes must precede effects, reversing the temporal order of a model should result in attenuating or reversing the sign of a prospective effect presumed to be causal. In other words, when time is reversed so that subsequent states predict prior states (e.g., $X_2 \rightarrow X_1$), the direction of lagged effects between constructs (e.g., $Y_1 \rightarrow X_2$) should be reversed or reduced in size. In contrast, contemporaneous effects, being insensitive to temporal ordering, should instead remain unchanged (Haufe et al., 2013). Adapting this approach, the reply in Sorjonen et al. (2025b) introduced a "time-reversed" REM and argued that because the effect of ASC on ACH remained positive, the standard (time-aligned) REM from Marsh et al. (2024) was not causal in nature. Yet, the implementation of this time-reversal testing raised several issues.

First, their time-reversed model (Fig. 1, A) did not reflect the REM in Marsh et al. (2024) (Fig. 1, B.1 and C.1): It replaced the lagged effect $ASC_T \rightarrow ACH_{T+1}$ by a contemporaneous effect $ASC_T \rightarrow ACH_T$, omitted the contemporaneous effect $ACH_T \rightarrow ASC_T$, and removed the autoregressive structure for ASC. In addition, this model reversed time only in relation to ACH, leaving ASC time-aligned; and it repeated the specification error of the STARTS model (Table 1) by allowing contemporaneous effects to start on wave 1. These numerous changes in the REM structure, as well as the uneven time-reversal (for ACH but not for ASC), alter the comparison of the time-aligned REM in Marsh et al. (2024) and the time-reversed REM in Sorjonen et al. (2025b). Furthermore, based on their time-reversed model, Sorjonen et al. (2025b) observed that the contemporaneous effect $ASC_T \rightarrow ACH_T$ remained significant and interpreted this as evidence against causality, despite the fact that contemporaneous effects are expected to remain unchanged in time-reversed models. In short, the time-reversal testing departed in important ways from the original method (Haufe et al., 2013).

The time-reversal testing is nonetheless valuable for our scientific debate. We therefore applied this test, this time in alignment with the method (Haufe et al., 2013). As reported in Fig. 1 (B.2, C.2), time-reversed REM variants were estimated

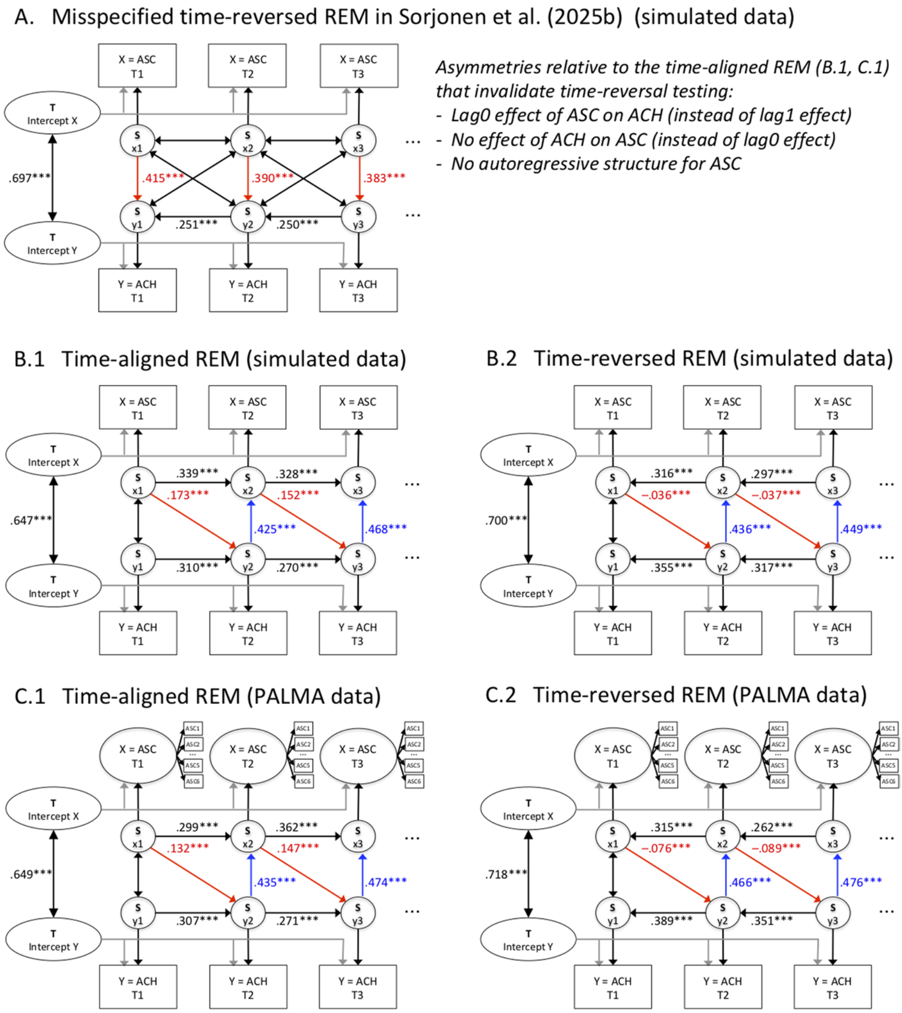

Fig. 1 Time-Aligned and Time-Reversed REM Across Simulated and PALMA Datasets. Note. ASC = academic self-concept; ACH = achievement. All models (except panel A) are based on the extended REM reported by Marsh et al. (2024), labeled "time-aligned REM" for short. In time-reversal tests, time-aligned and time-reversed REM are contrasted to exhibit directional consistency of lagged effects of ASC on ACH (Haufe et al., 2013). When ASC and ACH autoregressive structures are reversed (panels B.2 and C.2), effects of ASC on ACH are also reversed in sign, suggesting that their predictive effects exhibit directional asymmetry and reverse in direction when effects are allowed to precede causes (time-reversal). The time-reversed REM in Sorjonen et al. (panel A) introduces several changes to the REM from Marsh et al. (2024) that break its commensurability for time-inverse testing, while also contradicting SEM principles (e.g., no autoregressive structure, contemporaneous path starting on wave 1 instead of wave 2). ***$p < .001$

using the simulated data by Sorjonen et al. (2025a), as well as the original PALMA data from Marsh et al. (2024). Results showed patterns of effect changes consistent with the inversion of causal direction in time-reversal testing: The positive lagged effects of ASC on ACH identified in Marsh et al. became smaller and was reversed in sign (negative effects) in the time-reversed variants, while the contemporaneous effects of ACH on ASC remained stable (Fig. 1B.2, C.2). The present results thus show that the patterns reported in Marsh et al. (2024) change in ways expected under temporal inversion of the modeled relations, indicating that the observed pattern is consistent with the hypothesized directional ordering of ASC and ACH. These results are compatible with, but do not establish, a causal interpretation. Reproducible code for these analyses are available in our OSF repository (https://osf.io/ry6p3/overview?view_only=2af2bc8cd4694433b9024eb2577451b6).

Conclusion

The present commentary makes two distinct contributions. First, it clarifies modeling issues affecting the interpretability of alternative longitudinal models proposed by Sorjonen et al. (2025a, b). Second, it presents time-reversal analyses examining whether observed patterns are consistent with hypothesized temporal ordering. These contributions address complementary aspects of the debate and help advance a more rigorous understanding of how dynamic constructs may relate to each other over time.

In conclusion, the reply by Sorjonen et al. (2025b) revisits the debate on the REM, but does not fully engage with the central specification issues identified in their earlier models (Núñez-Regueiro et al., 2025). As argued throughout this response, these issues are quite technical and were not previously presented in sufficient detail, which may explain why they were overlooked in their reply. The present commentary therefore aimed to bridge this gap by providing a more explicit account of how certain modeling choices in Sorjonen et al.'s comments (Sorjonen et al. 2025a, b) depart from standard SEM logic and how these departures limit the interpretability of the results. In addition to these clarifications, we conducted renewed time-reversal tests showing that the reciprocal processes posited by the REM, are consistent with the patterns observed when time is reversed in the modeling strategy.

This commentary also highlights the importance of making modeling assumptions fully explicit when comparing alternative longitudinal models. Such clarification is necessary to ensure that differences in results reflect substantive mechanisms rather than specification issues in the models. Across this exchange, it becomes evident that what appeared to be contradictory results from alternative models (Sorjonen et al. 2025a, b) are better understood as consequences of specification issues that, once resolved, yield results consistent with the REM (Núñez-Regueiro et al., 2025).

Taken together, the juxtaposition of several stress tests, including alternative temporal structure and modeling strategies, and comparison of time-aligned and time-reversed models, contributes to a more precise understanding of the conditions under which reciprocal effects between academic self-concept and achievement can be meaningfully interpreted. Across these tests, the results reported by Marsh et al. (2024) and replicated in Núñez-Regueiro et al. (in press) remain consistent, showing that academic self-concept and achievement relate to each other over different timescales.

Author Contributions F.N.R: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing-Original Draft, Writing- Review and Editing, Visualization. H.M.: Conceptualization, Writing-Original Draft, Writing- Review and Editing. R.P.: Conceptualization, Writing- Review and Editing. O.L.: Conceptualization, Writing- Review and Editing. J.G.: Conceptualization, Writing- Review and Editing.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Data Availability Data for reproducing the time-reversal testing results are available on the Open Science Framework repository: https://osf.io/ry6p3.

Declarations

Conflict of Interest The authors report no conflict of interest.

References

Bailey, D. H., Jung, A. J., Beltz, A. M., Eronen, M. I., Gische, C., Hamaker, E. L., Kording, K. P., Lebel, C., Lindquist, M. A., Moeller, J., Razi, A., Rohrer, J. M., Zhang, B., & Murayama, K. (2024). Causal inference on human behaviour. *Nature Human Behaviour, 8*(8), 1448–1459. https://doi.org/10.1038/s41562-024-01939-z

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications.

Freund, P. A., & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin, 138*(2), 296–321. https://doi.org/10.1037/a0026556

Ghisletta, P., & McArdle, J. J. (2012). Latent curve models and latent change score models estimated in R. *Structural Equation Modeling: A Multidisciplinary Journal, 19*(4), 651–682. https://doi.org/10.1080/10705511.2012.713275

Haufe, S., Nikulin, V. V., Müller, K.-R., & Nolte, G. (2013). A critical assessment of connectivity measures for EEG data: A simulation study. *NeuroImage, 64*, 120–133. https://doi.org/10.1016/j.neuroimage.2012.09.036

Kenny, D. A., & McCoach, D. B. (2025). Ruling out latent time-varying confounders in two-variable multi-wave studies. *Multivariate Behavioral Research, 0*(0), 1–17. https://doi.org/10.1080/00273171.2025.2503829

Kenny, D. A., & Zautra, A. (1995). The trait-state-error model for multiwave data. *Journal of Consulting and Clinical Psychology, 63*(1), 52.

Kievit, R. A., Brandmaier, A. M., Ziegler, G., van Harmelen, A.-L., de Mooij, S. M. M., Moutoussis, M., Goodyer, I. M., Bullmore, E., Jones, P. B., Fonagy, P., Lindenberger, U., & Dolan, R. J. (2018). Developmental cognitive neuroscience using latent change score models: A tutorial and applications. *Developmental Cognitive Neuroscience, 33*, 99–117. https://doi.org/10.1016/j.dcn.2017.11.007

Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th ed.). Guilford Publications.

Marsh, H. W., Guo, J., Pekrun, R., Lüdtke, O., & Núñez-Regueiro, F. (2024). Cracking chicken-egg conundrums: Juxtaposing contemporaneous and lagged reciprocal effects models of academic self-concept and achievement's directional ordering. *Educational Psychology Review, 36*(2), Article 53. https://doi.org/10.1007/s10648-024-09887-w

Muthén, B., & Asparouhov, T. (2024). Can cross-lagged panel modeling be relied on to establish cross-lagged effects? The case of contemporaneous and reciprocal effects. *Psychological Methods*. https://doi.org/10.1037/met0000661

Núñez-Regueiro, F., Marsh, H. W., Bressoux, P., Batruch, A., Bouet, M., Bressan, M., Brown, G., Butera, F., Cherbonnier, A., Darnon, C., Demolliens, M., de Place, A. L., Desrichard, O., Goron, L., Hémon, B., Huguet, P., Jamet, E., Mazenod, V., Mella-Barraco, N., & Pansu, P. (2026). I feel competent, therefore I am: Self-concept and skill interact at different speeds. *Educational Psychology Review, 38*(39).

Núñez-Regueiro, F., Marsh, H. W., Pekrun, R., Lüdtke, O., & Guo, J. (2025). The reciprocal effects model is robust to alternative modeling specifications: A response to Sorjonen et al., 2025. *Educational Psychology Review, 37*(3), Article 79. https://doi.org/10.1007/s10648-025-10059-7

Sorjonen, K., Ingre, M., Nilsonne, G., & Melin, B. (2023). Dangers of including outcome at baseline as a covariate in latent change score models: Results from simulations and empirical re-analyses. *Heliyon*. https://doi.org/10.1016/j.heliyon.2023.e15746

Sorjonen, K., Melin, B., & Nilsonne, G. (2025a). Inconclusive evidence for a prospective effect of academic self-concept on achievement: A simulated reanalysis and comment on Marsh et al. (2024). *Educational Psychology Review, 37*(2), Article 30. https://doi.org/10.1007/s10648-025-10008-4

Sorjonen, K., Melin, B., & Nilsonne, G. (2025b). Still no convincing evidence for a causal effect of academic self-concept on achievement: A reply to Núñez-Regueiro et al., 2025. *Educational Psychology Review*, *37*(4), 98. https://doi.org/10.1007/s10648-025-10083-7

Sorjonen, K., Melin, B., & Nilsonne, G. (2025c). The Spurious Prospective Associations Model (SPAM): Explaining longitudinal associations due to statistical artifacts. *PLOS ONE, 20*(9), Article e0331609. https://doi.org/10.1371/journal.pone.0331609

Usami, S., Hayes, T., & McArdle, J. J. (2015). On the mathematical relationship between latent change score and autoregressive cross-lagged factor approaches: Cautions for inferring causal relationship between variables. *Multivariate Behavioral Research*, *50*(6), 676–687. https://doi.org/10.1080/00273171.2015.1079696

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, *24*(5), 637–657. https://doi.org/10.1037/met0000210

Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K. A. Bollen, & J. S. Long (Eds.), *Testing Structural Equation Models*. Sage Publications, Inc.

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science, 9*(2), 111–125. https://doi.org/10.1177/1745691613518075

Authors and Affiliations

Fernando Núñez-Regueiro[1] · **Herbert W. Marsh**[2,3] · **Reinhard Pekrun**[2,4] · **Oliver Lüdtke**[5,6] · **Jiesi Guo**[2]

✉ Herbert W. Marsh
herb.marsh@acu.edu.au

Fernando Núñez-Regueiro
fernando.nr.france@gmail.com

Reinhard Pekrun
reinhard.pekrun@acu.edu.au

Oliver Lüdtke
oluedtke@leibniz-ipn.de

Jiesi Guo
jiesi.guo@acu.edu.au

1 Contextual Learning Research Laboratory (LaRAC), Université Grenoble Alpes, Grenoble 38000, France

2 Institute for Positive Psychology and Education, Australian Catholic University, North Sydney 2060, Australia

3 Oxford University, Oxford, UK

4 University of Essex, Colchester, UK

5 Department of Educational Measurement, IPN – Leibniz Institute for Science and Mathematics Education, Olshausenstraße 62, Kiel 24118, Germany

6 Centre for International Student Assessment (ZIB), Kiel, Germany