Identifying Reciprocities in School Motivation Research: A Review of Issues and Solutions Associated With Cross-Lagged Effects Models

Fernando Núñez-Regueiro, Jacques Juhel, Pascal Bressoux, Cécile Nurra

Identifying Reciprocities in School Motivation Research: A Review of Issues and Solutions Associated With Cross-Lagged Effects Models

Fernando Núñez-Regueiro[1], Jacques Juhel[2], Pascal Bressoux[1] and Cécile Nurra[1]

[1]Université Grenoble Alpes

[2]Université Rennes 2

Author Note

This research is based on data also used in the doctoral dissertation of the first author and in other manuscripts submitted for publication. The data collection was supported by a doctoral grant from Appel à projets Grenoble Innovation Recherche ("DESCOL") to Fernando Núñez-Regueiro.

The data are available at https://osf.io/mrfw4

Correspondence concerning this article should be addressed to Fernando Núñez-Regueiro, Department of Education, Université Grenoble Alpes, Laboratoire de Recherche sur les Apprentissages en Contexte, 1251 Avenue Centrale, BP 47, 38040 Grenoble Cedex 9, France. Email: fernando.nunez-regueiro@univ-grenoble-alpes.fr

Abstract

Part of the evidence used to corroborate school motivation theories relies on modeling methods that estimate cross-lagged effects between constructs, that is, reciprocal effects from one occasion to another. Yet, the reliability of cross-lagged models rests on the assumption that students do not differ in their trajectories of growth over time (e.g., no high- or low-achievers). The present review explains why deviations from this assumption produce unreliable findings by confounding between- and within-person processes of change. To relax this assumption, next-generation cross-lagged models are presented and illustrated using panel data on high school students ($N = 944$). These issues and solutions are discussed using, as a case study, the pervading theory that motivation develops as a function of reciprocal effects between beliefs about the self (e.g., academic self-concept) and school achievement. Implications regarding the use of cross-lagged models and knowledge building in school motivation research are discussed. Supplementary materials containing technical notes on cross-lagged models, as well as open-source data and scripts for R and Mplus, are provided to aid educational researchers use and compare these alternative models.

Keywords: cross-lagged effects, review, school motivation, Granger causality, convergence assumption

Supplemental Materials: https://doi.org/10.1037/edu0000700.supp

Educational Impact and Implications Statement

Current school motivation theories generally agree that students' academic self-beliefs (e.g., feelings of competency or efficacy as a student) and school performance (e.g., grade point average, learning acquisitions) reinforce each other over time, meaning that experiencing positive developments on one dimension causes positive developments on the other dimension as well. However, the evidence upon which this conclusion has been reached in the literature relies on conventional modeling techniques that are prone to confound changes between individuals (i.e., how students differ on these dimensions over time, relative to one another) and changes within individuals (i.e., how students evolve on these dimensions over time, relative to themselves). In this article, we explain under which conditions this confounding can lead to erroneous claims about reciprocal influences, and how novel modeling techniques can fix this issue and produce more reliable evidence. Based on a literature review and on original data, evidence is reported that the above theoretical claim (i.e., that academic self-perceptions and school performance mutually influence each other) is vulnerable to such confounding and does not stand to the scrutiny of novel modeling techniques. Complementary research might therefore be conducted to investigate this issue further and uncover novel findings concerning processes of change in student motivation.

"*5.6. The limits of my language mean the limits of my world.*"

(Wittgenstein, 2021/1921, p.68)

Identifying Reciprocities in School Motivation Research

A central endeavor of school motivation theories is to understand "how and why motivation changes over time and how motivational processes lead to change in students' social and academic competencies" (Wentzel & Wigfield, 2009, p. 7). One of the challenges associated with this research concerns the selection of modeling strategies that might provide evidence for the processes of change depicted in motivational theories. Arguably, the theoretical depiction of developmental processes is influenced by the modeling capabilities themselves and, more precisely, by the conceptual tools they provide mathematically to comprehend processes of change. By implication, the language afforded by the modeling strategies might also limit one's capability to describe processes of change in an appropriate manner (theoretically). Aligning with this view, the present article argues that popular modeling strategies (i.e., residualized change model, cross-lagged effect panel model, autoregressive model) have limited school motivation theories' capability of describing developmental processes, by providing conceptual tools that conflate processes of change situated within- versus between- individuals. We argue that, when the latter processes of change differ, standard models offer an inadequate representation of reciprocal effects between constructs over time, both epistemologically (ecological fallacy) and empirically (statistical artifacts). To shed light on these limitations, a new set of conceptual tools is presented in the form of "next-generation cross-lagged effects models" (Curran et al., 2014; Hamaker et al., 2015; Zyphur, Allison, et al., 2020), which enable disaggregating within- and between-person processes of change.

The innovations provided by next-generation models are quite recent and have not been

sufficiently integrated or discussed within the field of school motivation. Accumulating evidence in other fields has now demonstrated that standard cross-lagged effects models are prone to produce statistical artifacts that lead to erroneous claims concerning reciprocities between dynamic constructs (e.g., truly non-significant cross-lagged effects appearing as significant, or conversely), which are made more apparent in next-generation models (Berry & Willoughby, 2017; Hamaker et al., 2015; Mund & Nestler, 2019; Zyphur, Allison, et al., 2020). However, although next-generation models are gaining increasing attention in the school motivation research, rare applications in the field have only used a single kind of next-generation model (i.e., the random-intercept cross-lagged effect model) that fails to disaggregate between- and within-person processes of change if individuals differ in the shape of their developmental trajectories (Brandt et al., 2019; Burns et al., 2019; Ehm et al., 2019; Ruzek & Schenke, 2019). Moreover, despite their limitations, standard models are still widely used across empirical studies and meta-analyses published in major journals in the field (Arens et al., 2017; Cai et al., 2019; Engels et al., 2019; Garon-Carrier et al., 2016; Quin, 2017; Sutter-Brandenberger et al., 2018; Vasconcellos et al., 2019; Weidinger et al., 2017). These uncertain research practices can be explained by the fact that next-generation methods have been published outside the field of education, in technical introductions hardly accessible to applied researchers (Curran et al., 2014; Hamaker et al., 2015; Zyphur, Allison, et al., 2020), and with very little discussion of their similarities and differences (Mund & Nestler, 2019; Usami et al., 2019). In this context, educational psychologists might be in want of clarification concerning the meaning and added value of these methods. Should we really embrace next-generation cross-lagged models in place of standard cross-lagged models, and for what reasons?

The present article aims at answering this question by reviewing, within the context of

school motivation research, the issues and solutions of statistical models aimed at identifying reciprocal causal effects in non-experimental designs, notably standard (i.e., cross-lagged panel model, residualized change model; Little et al., 2007; Selig & Little, 2012) and next-generation cross-lagged effects models (Curran et al., 2014; Hamaker et al., 2015; Zyphur, Allison, et al., 2020). First, we present the "reciprocal effects model" (REM), a pervading school motivation theory according to which self-beliefs and achievement reinforce each other over time, to create positive or negative cycles of individual development (Marsh & Craven, 1997, cited in Marsh & Craven, 2006). Second, we use this theory as a case study to illustrate the epistemological and methodological issues associated with standard cross-lagged effects models, and the solutions offered by next-generation models. Third, we review studies that have compared findings from both modeling strategies, to show the added value of next-generation models in the investigation of reciprocal effects, including REM processes. Finally, an empirical section tests the REM theory under these various modeling strategies and analyzes their discrepancies. By outlining the strengths and weaknesses of standard and next-generation cross-lagged effects models, our goal is to shed new light on their use in empirical studies and on the state of knowledge in school motivation theories.

Literature Review on the Reciprocal Effects Model

Educational theories generally agree that self-beliefs—such as beliefs about one's worth, capabilities, adequacy at school—are motivational mediators of the relationship between educational contexts and student achievement (Christenson et al., 2012; Wentzel & Wigfield, 2009). This view is condensed eloquently in the REM theory of reciprocal effects between school achievement and self-beliefs (Marsh & Craven, 1997, cited in Marsh & Craven, 2006). REM processes are among the most consensual in the field of school motivation and offer a

general case study for the way reciprocities are analyzed in the field.

Theories and Evidence on REM Processes

The idea of reciprocal effects between self-beliefs and school achievement has been a recurring theme in school motivation theories. Indeed, theories based on self-efficacy (Pajares & Schunk, 2001), self-concept (Marsh et al., 2019), self-determination (Pitzer & Skinner, 2016; Reeve, 2012), self-regulation (Panadero, 2017), or achievement emotion processes (Pekrun, 2006; Pekrun & Linnenbrink-Garcia, 2012) all recognize that believing in one's abilities, competences or capabilities at school motivates students to engage and persevere in learning activities which, in turn, improves school performances (self-enhancement hypothesis). Moreover, these models also assume that prior school performances are positively related to later self-beliefs, either directly or indirectly by fostering supportive social relationships and motivational or meta-cognitive processes (skill-development hypothesis; Table 1). Complementary examples of overlap between the REM and other motivational theories (e.g., self-affirmation theory, self-worth theory, possible selves theory) have been presented elsewhere (for details, see Valentine et al., 2004). Psychological theories therefore share a common ground in REM processes and only differ regarding the intermediate processes between self-beliefs and achievement (e.g., dimensions of engagement or motivation, focus on context or self-regulation) or regarding the kinds of self-beliefs they privilege (as discussed hereafter). Conversely, due to its neutrality vis-à-vis intermediate processes, the REM is compatible with multiple psychosocial explanations (Valentine et al., 2004). In this sense, it offers a powerful benchmark to evaluate the empirical validity of most school motivation theories.

—Insert Table 1 here—

Although it appears as a keystone, the evidence for the REM has only recently been

recognized as unequivocal and robust. Indeed, scholars have long debated over whether self-beliefs cause achievement or whether achievement causes self-beliefs (Calsyn & Kenny, 1977, cited in Marsh & Craven, 2006; Pajares & Schunk, 2001), but literature reviews and meta-analyses in the last decade have reported evidence for the existence of a bidirectional causality. More precisely, Valentine and colleagues (2004) and Valentine (2001, cited in Huang, 2011) reviewed 55 studies published between 1951 and 1996 representing 60 independent samples and 282 effects sizes. Their meta-analyses estimated that, controlling for prior levels and other influential social and demographic variables, the mean effect of prior self-beliefs (i.e., self-concept, self-efficacy, self-esteem) on later achievement is .08, whereas the mean reciprocal effect is .16 (standardized coefficients). Based on 32 studies published between 1976 and 2007 and correcting for measurement error, Huang (2011) estimated these reciprocal effects to be both equal to .12. Meta-analyses therefore concur that reciprocal effects between self-beliefs and achievement do exist. In addition, REM processes appear to generalize to various school contexts (e.g., academic vs. vocational track; Arens et al., 2017), to other domains than school (e.g., health, sports) and to various national contexts (e.g., United States, Germany, Canada, Hong Kong, China), which also points to their robustness (Marsh & Craven, 2006; Marsh & Martin, 2011; Valentine et al., 2004).

REM Processes From a Multidimensional Point of View

Theories and evidence support a multidimensional approach on self-beliefs. According to the literature, self-esteem represents most of the time a general construct that values overall abilities from an affective perspective (i.e., self-worth as a person; Harter, 1999; Marsh, 1986; Marsh & O'Mara, 2008); self-concept is usually a construct sharing both descriptive (i.e., perceived capacities) and evaluative facets (i.e., feelings induced by internal or social

comparison processes) that is retrospective in nature (i.e., based upon past experiences) and generally defined on an intermediate level of specificity (feelings of competency at school), but that can also be measured specifically (i.e., feelings of competency in a given discipline; Bong & Skaalvik, 2003; Harter, 1999; Marsh et al., 2019; Marsh & Craven, 2006; Pajares & Schunk, 2001; Skinner, 1996); and self-efficacy is a task-oriented construct, usually descriptive and prospective in nature, that is usually measured specifically (i.e., capability to do a set of mathematical tasks in the future; Bandura, 1986, 1997; Bong & Skaalvik, 2003; Marsh et al., 2019; Pajares & Schunk, 2001). The REM research has primarily focused on academic self-concept and, secondarily, on self-efficacy and self-esteem (Huang, 2011; Marsh & Craven, 2006; Pajares & Schunk, 2001; Valentine et al., 2004). Meta-analyses find that REM processes are strongest at specific or intermediate levels of analysis (vs. general) and that no significant differences between self-beliefs are found once level of specificity is accounted for (Huang, 2011; Valentine et al., 2004).

In sum, the reciprocal effects model between self-beliefs and achievement has received support from various meta-analyses and literature reviews and across multiple settings. Evidence also points to the importance of a multidimensional approach on REM processes that distinguishes self-beliefs by level of specificity, with the core of the theory focused on an intermediate level, that is, academic self-concept. More generally, this model is an important contribution to the literature because it rests on apparently solid evidence and because it gives predictive validity to most theories in the field of educational psychology, notably on bidirectional causalities that are seldom investigated otherwise.

A Critique of Standard Cross-Lagged Effects Methods in REM Research: From Ecological Fallacies to Statistical Artifacts

In designs with randomized experimental groups, the causal effect of a variable on an outcome can be investigated in the form of a treatment effect, by manipulating the variable across experimental groups and estimating differences in outcomes between groups (i.e., Rubin's causal model; Holland, 1986). Econometric techniques can be used to obtain similar conditions in quasi-experimental designs (Imbens & Wooldridge, 2009). In non-experimental designs or designs not amenable to this framework, an alternative form of effect can be investigated to provide tentative evidence of causality, by estimating the predictive effect of a variable on an outcome variable from one occasion *T* to a later occasion *T+1*, also known as a "lagged effect" (i.e., Granger's causal model; Holland, 1986). To ensure that lagged effects are not confounded with other sources of variance, the Granger causal model assumes "(a) a cause→effect temporal order, (b) possible bidirectional effects among variables, and (c) controls for potential confounders" (Zyphur, Voelkle, et al., 2020, p. 2). In sum, educational researchers seeking to identify causal effects between two constructs, or cross-lagged effects, will resort to Rubin's causal framework in experimental settings and to Granger's causal framework in non-experimental settings (for other putative causal frameworks, see "Discussion").

Because ecological settings such as schools seldom allow manipulating students' motivational processes, educational psychologists willing to test bidirectional causalities between achievement and self-beliefs (REM processes) often adhere to Granger's causal framework and seek to identify cross-lagged effects. In this perspective, two modeling methods have been in use, namely the "residualized change" model (RC) and the "cross-lagged panel model" (CLPM). The RC model uses longitudinal data and controls for initial individual differences (i.e.,

including covariates and baseline values of the outcome variable) to meet, respectively, conditions (a) and (c) of the Granger causality. The CLPM and its variants (full-forward model, autoregressive model) augment the RC by specifying covariances, autoregressive effects and reciprocal effects between two or more constructs of scientific interest, thus apparently meeting condition (b) of the Granger causality (Bollen, 1989; Selig & Little, 2012). Because the CLPM appears to meet all conditions to identify (Granger) causal effects, REM researchers have considered it as the state-of-the-art method to test REM processes in ecological settings (Marsh & Craven, 2006). However, recent methodological developments show that the CLPM (and, by implication, the RC) might produce unreliable estimates by confounding processes of change situated between versus within individuals (Berry & Willoughby, 2017; Curran et al., 2014; Hamaker et al., 2015; Zyphur, Allison, et al., 2020). Importantly, these developments shed light on the "ecological fallacies" by which lower-level processes are inferred from evidence on higher-level processes (Curran & Bauer, 2011; Robinson, 1950), an epistemological issue seldom discussed in applied research. After a discussion of two assumptions that undergird these ecological fallacies, we then explain the modeling issues that result from the conflation of within- and between-person processes of change, as illustrated in the REM research.

A Gap Between Theory and Evidence on REM Processes: On Ecological Fallacies Due to Non-Convergence

From an epistemological point of view, there appears to be a gap between the processes of change in self-beliefs and achievement as portrayed in the REM theory, that is, *changes within the individual* (Marsh & Craven, 2006; Marsh & Martin, 2011); and the processes of change identified in empirical studies of the REM by using the CLPM, which confound *changes within and between individuals over time* (Curran et al., 2014; Hamaker et al., 2015). This gap turns out

to be problematic when within-person and between-person processes of change are not equivalent in their dynamics, in which case ignoring this gap in knowledge building constitutes an ecological fallacy (Berry & Willoughby, 2017; Curran et al., 2014; Robinson, 1950).

But to what extent can we say that the REM research is vulnerable to ecological fallacy? The answer to this question depends on two assumptions that are well defined in the statistical literature but are seldom recognized in applied research—and apparently amalgamated in some discussions (e.g., Curran & Bauer, 2011; Curran & Wirth, 2004; Selig & Little, 2012)—, namely the assumptions of convergence and ergodicity. The assumption of convergence (Berry & Willoughby, 2017; Hoffman, 2015) states that the changes in the differences between individuals over time are equivalent to the changes over time within individuals or, equivalently, that no differences between individuals' developmental trajectories exist. As explained in greater detail hereafter, violating this assumption will lead to a distortion of the parameters of cross-lagged effects and, thus, to erroneous conclusions about reciprocal effects. The assumption of ergodicity further assumes that individuals obey the same dynamical system and are therefore interchangeable (homogeneity assumption), which means that the average dynamical system (demographic structure) is equivalent to each individual dynamical system (idiographic structure); and that the process of change is stationary (stationarity assumption), which means that the function describing this process is invariant over time and can therefore be estimated from the changes incurred during this period (Molenaar, 2004, 2008)[1]. When ergodicity is

[1] More precisely, for Gaussian time series, stationarity is the property of a dynamical system for which the mean, variance and lagged covariance structure are invariant in time (Molenaar, 2004). A linear dynamical system with an unstable trajectory of growth (as those observed in REM processes) can also be transformed to stationarity after including a model for its trajectory, outside the autoregressive part of the model (Asparouhov et al., 2018). Stationarity in this sense is a necessary condition for pooling multiple measurement occasions together into a single functional form (Molenaar, 2004). However, panel data can be non-stationary in different ways (Hamaker, 2021). The "stationarity" hypotheses more specifically tested in the CLPM only assume that some or all of the aggregated

compromised, the parameters obtained from cross-lagged analysis do not reflect individual development anymore and are thus hardly interpretable, even if convergence is found to be true. On the contrary, when both assumptions are met (i.e., convergence and ergodicity), the population parameters that describe the mean differences between individuals over time provide valid evidence to infer on the idiosyncratic dynamics that describe the developmental system of each individual, which is what is afforded by the CLPM and other commonly used identification strategies (residualized change analysis, path analysis, etc.).

To further illustrate the central assumptions made by the empirical models currently used to validate the REM, Figure 1 portrays two individuals *P* and *Q* experiencing variations in their grades over time, in what is assumed to be a stationary process (ergodicity is therefore assumed when homogeneous dynamics are observed). Two series of graphs [a) to c) and d) to f)] are presented. In both series, the first graph represents the dynamics of each individual as being qualitatively or "genetically" different (non-homogeneity) and as presenting differences between developmental trajectories [non-convergence; Figure 1, a) and d)]. Some modeling methods allow this specification (for an example, see "Discussion"). The second graph represents individual dynamics as showing quantitative differences in trajectories as well (non-convergence), but a common dynamical system *(P,Q)* is assumed to prevail [homogeneity; Figure 1, b) and e)]. This may be specified in the next-generation cross-lagged panel models presented hereafter (see subsection "Relaxing the Convergence Assumption"; Curran et al.,

---

parameters for the population of dynamical systems (e.g., cross-lagged effects, autoregressive effects) are equivalent in size across measurements for equally spaced observations (Little et al., 2007; Selig & Little, 2012). In fact, the CLPM assumes dynamical stationarity and homogeneity as prerequisites to obtain meaningful model parameters and, therefore, to test parametric stationarity hypotheses (as defined in the CLPM). By contrast, the condition of stationarity in CLPM parameters is not sufficient to provide evidence for stationarity in the dynamical systems undergirding processes of change.

2014; Hamaker et al., 2015; Zyphur, Allison, et al., 2020). The third graph adheres with both assumptions by considering that individuals are interchangeable in their dynamics (homogeneity) and present no differences in their trajectories [Figure 1, c) and f)], which corresponds to the CLPM specification. In each case, similar assumptions can be extended with regards to the dynamics of self-beliefs.

—Insert Figure 1 here—

In substance, testing the REM with the CLPM can provide valid evidence if we assume that students' self-beliefs and achievement are driven by a stationary process and by common structural dynamics (homogeneity leading to ergodicity), and if we further assume that students show no developmental differences in either construct [convergence; Figure 1, c) and f)]. To date, the REM theorists have not postulated differences in the dynamical systems undergirding achievement and self-beliefs, nor have they postulated bifurcations due to non-stationary processes (Arens et al., 2017; Marsh & Craven, 2006). Indeed, the positive or negative developmental loops described in the REM appear as two opposite versions of a common dynamical system that unfolds linearly over time. This means that the REM theory assumes ergodicity, albeit implicity. Moreover, the REM theory states that, within each individual, "academic self-concept and achievement are mutually reinforcing, each leading to gains in the other" (Marsh & Craven, 2006, p. 133). This definition of REM processes is ambiguous with regards to its theory of change, because it does not state whether such changes are temporary within-person processes ("intraindividual variability") or enduring between-person processes ("intraindividual change"; Nesselroade, 1991). We may assume that the REM theory views both kinds of changes as equivalent and that it focuses on within-person processes. However, this view stands in contrast with some of the most consensual observations in the literature, namely

that students differ in terms of their growth trajectories of achievement and self-beliefs at school (Fraine et al., 2007; Nagy et al., 2010; Parker et al., 2015; Pfost et al., 2014). The latter observations reject the convergence hypothesis. For this reason, the REM theory is incompatible with a CLPM specification that assumes convergence. By implication, from a strictly epistemological point of view, the REM is vulnerable to the ecological fallacy of inferring evidence on within-person processes of change from evidence partly based on between-person processes of change, when in fact the two kinds of processes differ (non-convergence).

Methodological Issues in REM Research: On Statistical Artifacts Due To Non-Convergence

Over and above the epistemological problem of an ecological fallacy, the incompatibility between standard lagged effects models (CLPM, residualized change, path analysis, etc.) and the convergence assumption (i.e., assuming that individuals do not differ in their developmental trajectories) poses a methodological problem that undermines the interpretation of model parameters. Indeed, recent developments demonstrate that, when the convergence assumption does not hold, the parameters identified in lagged effects models represent an amalgam of between-person and within-person processes that generates spurious findings (Berry & Willoughby, 2017; Hamaker et al., 2015; Mund & Nestler, 2019; Zyphur, Allison, et al., 2020). Moreover, even in cases where between- and within-person processes actually converge, this convergence is very difficult to obtain in practice due to its strong sensitivity to covariates included in the model (Hoffman, 2015). This calls into question the prevailing use of standard lagged effects models in school motivation research. To better appreciate this point, we focus in this section on revealing the spurious nature of the REM evidence based on the CLPM (cf. Figure 2) when the convergence assumption does not hold.

Let us consider the schematic representations of Figure 1 and, in particular, the case

where the homogeneity assumption holds and where the convergence assumption does not hold due to stable differences between individual trajectories. In this case (Figure 1, b), mean differences between individuals (between-person variance) overshadow the variable differences around each individual's own mean (within-person variance). Because the within-person variance is small relative to the between-person variance, the estimated lagged-effect is mostly determined by the between-person differences and, thus, by the fact that, on average, individual *P* always fares better than individual *Q*. In this configuration, a standard lagged effects analysis will conclude that having higher-than-average grades (relative to the average individual) on one occasion induces gains in grades on another occasion (i.e., positive autoregressive effect β_4 in Figure 2). Likewise, if the processes of change in self-beliefs show similar patterns as grades—with individual *P* showing a higher mean self-belief than individual *Q*, and each individual showing variations around its own mean—, the analysis will conclude that having higher-than-average self-beliefs on one occasion induces gains in self-beliefs on another occasion (i.e., positive autoregressive β_1 in Figure 2). Moreover, because of variability around the mean on each construct (i.e., within-person variance), the autoregressive effects for each construct will be estimated as being high but partial rather than complete (i.e., standardized coefficient < 1.000), thus freeing unexplained variations in gains that will be picked up by the cross-lagged effects of one construct onto another (coefficients β_2 and β_3 in Figure 2). Again, these cross-lagged effects will nevertheless reflect the mean individual differences (*P* > Q) if the variability around the mean (within-person variance) is small relative to the differences between means (between-person variance). These observations will not be very sensitive to controls of the covariance between residuals in grades and self-beliefs on each occasion (covariance between residuals ε in Figure 2), because these pick up the within-person variance not accounted for by the mean

between-person differences that really drive the autoregressive and cross-lagged effects.

—Insert Figure 2 here—

In terms of the CLPM (cf. Figure 2), the above explanations mean that the autoregressive effects β_1 and β_4 will likely be positive (due to the prevailing effect of mean positive differences) and strong albeit inferior to 1 (due to low within-person variance), and that the cross-lagged effects β_2 and $\beta_3$will likely be positive (due to the prevailing effect of mean positive differences) and low (due to the low magnitude of the within-variance they pick up). Yet, all these observations will be due to stable individual differences, not by actual lagged effects of self-beliefs on grades or of grades on self-beliefs. In other words, the identified effects will be spurious. Similar effects are observed across REM studies, with meta-analyses reporting, on average, high autoregressive effects (between .52 and .70) and low cross-lagged effects (between .08 and .16; Huang, 2011; Valentine et al., 2004). As explained hereafter, one empirical study has indeed demonstrated the existence of spurious effects in a CLPM test of the REM model, by showing that the cross-lagged effects were only due to the omission of stable differences between students' trajectories (Ehm et al., 2019). Importantly, as shown in Figure 1 e), individual trajectories can also differ as a function of growth factors (e.g., slopes), which can create non-convergence over and above the effects of stable differences in intercepts.

Relaxing the Convergence Assumption: The Case for Next-Generation Cross-Lagged Panel Models

To overcome the potential ecological fallacies and the identification issues of standard lagged effects models such as the CLPM, researchers might consider using next-generation cross-lagged panel models that relax the convergence assumption, notably by controlling for differences between individuals' growth trajectories (Curran et al., 2014; Hamaker et al., 2015;

Zyphur, Allison, et al., 2020). By cancelling out the between-person variance associated with growth trajectories, next-generation models enable the identification of lagged effects using unequivocally within-person variance, thus revealing dynamics previously overshadowed (cf. Figure 1). For instance, the relative difference between *P* and *Q* which appeared to be always positive in the non-convergence setting (Figure 1, b, e) now becomes negative on occasions T1 and T2 (*P* < Q), null on occasion T3 (*P* = Q) and positive on occasion T4 (*P* > Q; Figure 1, c, f). To date, two kinds of next-generation models exist that enable relaxing the convergence assumption. This section explains their specification and interpretation relative to the REM theory, in a view to illustrate and facilitate their use in educational research. Table 2 offers an overview of these models and their variants (for a mathematical presentation of these models, see Supplementary Material A).

The Random Effects Cross-Lagged Panel Model (RE-CLPM)

The first kind of next-generation cross-lagged model can be conceived as a combination of the CLPM (Selig & Little, 2012) and the bivariate growth curve model (Bollen & Curran, 2006). In this combination, the bivariate growth curve model aims at picking up all the variance that relates to individual differences in the growth trajectories of the constructs of interest (i.e., grades and self-beliefs in the REM theory), notably by specifying individually-varying or "random" effects. We shall therefore unite these models under the label "random effects cross-lagged panel models" (RE-CLPM). When individuals do not differ in their trajectories of growth (as in Figures 1, c or 1, g), the growth curve model is null and the RE-CLPM reduces to the CLPM. When individuals differ in their grades by a constant (as in Figure 1, b), the bivariate model picks up mean differences by allowing grades to be predicted by a person-specific constant or, stated otherwise, by a random intercept ("random intercept CLPM"; Hamaker et al.,

2015, p. 103). This might be the case, for example, when individuals are surveyed during a short period of time (e.g., over a period of days or a few weeks), so that individual differences remain largely stable. However, over longer periods of time as those investigated in REM research (e.g; over multiple school years; Arens et al., 2017; Marsh & Craven, 2006), individual differences will likely evolve as a function of other growth factors (as in Figure 1, e; Fraine et al., 2007; Nagy et al., 2010; Parker et al., 2015; Pfost et al., 2014). When individuals differ in their grades by the shape of their trajectory (e.g., increasing versus decreasing, U-shaped versus inverted-U-shaped), the bivariate model picks up the between-person variance by allowing grades to be predicted by a random intercept and additional individually-varying growth factors (e.g., linear and quadratic effects of time), as enabled in the "latent curve model with structured residuals" (LCM-SR; Curran et al., 2014, p. 879). For terminological and conceptual coherence with the random intercept CLPM (RI-CLPM), we shall call the latter model the "random curve CLPM" (RC-CLPM).

In order to obtain a proper control of these between-person differences, the bivariate growth curve model adjusts to each individual mean trajectory, by including all significant growth factors that contribute to differentiate individual trajectories. For instance, Figure 3 shows an application of the RC-CLPM where grades and self-beliefs are determined by an intercept and a slope, but the data might suggest otherwise in other cases (e.g., non-significant slope of self-beliefs). In terms of model building, authors recommend identifying random effects first, before testing CLPM parameters (i.e., autoregressive and cross-lagged effects) and covariates of random effects (Curran et al., 2014). A likelihood ratio test can also be conducted to compare the RE-CLPM with a reduced version including CLPM parameters, but without the random effects—that is, the CLPM which is nested in the RE-CLPM (Hamaker et al., 2015). It is

also important that the model controls for the covariance between growth factors within and across constructs (covariances ψ in Figure 3), to account for the possibility that the developmental differences on one construct (e.g., the intercept of grades) covary with the developmental differences in the evolution of the same construct (e.g., the slope of grades) or with another construct (e.g., the intercept or slope of self-beliefs). This prevents confounding the latter covariance with within-person variance. Thus, covariances between growth factors should be tested as well, but they are sometimes overlooked in applications of the RC-CLPM (e.g., Berry & Willoughby, 2017; Mund & Nestler, 2019). Note that the identification of random effects will require at least three waves of data for the RI-CLPM, and a least four waves of data for the RC-CLPM (Mund & Nestler, 2019).

—Insert Figure 3 here—

Once the growth parameters and their covariances are identified, the lagged effects of the RE-CLPM variants (i.e., either RI-CLPM or RC-CLPM) can be estimated from exclusively within-person variance, represented by the residuals ε_{S-B} and ε_{AC} in Figure 3. Technically, these residuals represent the variability in self-beliefs and school achievement on each occasion not accounted for by mean between-person differences (i.e., the growth factors). In substance, this kind of residuals can be conceived as an "impulse" or shock to the dynamical system that causes a deviation from the expected trajectory of growth. In this perspective, the autoregressive effects of residuals within constructs in the RC-CLPM (β_1 and β_4 in Figure 3) can be interpreted as representing the persistence of the effect of an impulse from one occasion to another. In extreme cases, a strong impulse might persist durably and produce a strong autoregressive effect within the dynamical system, possibly changing its nature (i.e., qualitative change or, in dynamical systems terms, "bifurcation"; Molenaar, 2001, p. 15425). Being person-specific, such an

impulse would break up the similarity between dynamical systems and compromise the assumption of homogeneity (cf. Figure 1, a and d). A RC-CLPM specification might be hardly compatible with such an impulse. On the contrary, if the impulse is mild enough, the persistence of its effect is likely to be short-run and the autoregressive effects will be small as well, thus allowing the initial dynamical system to regain stability in the long-run (stationarity assumption of the RC-CLPM). Following a similar rationale, the cross-lagged effects of the RC-CLPM (β_2 and β_3 in Figure 3) can be conceived as the impact of an impulse (on occasion *T*) on a construct in a subsequent occasion (e.g., grades on occasion *T+1*) that transits through its impact on another construct (e.g., self-belief on occasion *T*). Because the joint impact of the impulse on both constructs (on occasion *T*) is controlled for by the covariance between residuals, the cross-lagged effects are specifically related to the changes in one construct that explain the subsequent changes in another construct. Importantly, to the extent that individuals are considered interchangeable in their dynamics (homogeneity assumption), each of the person-specific impulses can be considered as a single experimental treatment that modifies a predictor to a certain degree (grades or self-beliefs) to assess its causal impact on a dependent variable (self-beliefs or grades; Zyphur, Voelkle, et al., 2020). In sum, to the extent that the REM theory assumes ergodicity, the RC-CLPM offers a powerful method to test the hypotheses that gains in self-beliefs (achievement) on an occasion cause gains in achievement (self-beliefs) on a later occasion, while relaxing the convergence assumption. For overviews comparing this first kind of next-generation models with other next-generation models that do not enable relaxing the convergence assumption (e.g., autoregressive latent trajectory model, factor CLPM, stable trait autoregressive and state model, dual latent change score model), see the works of Usami and collaborators (2019) and Mund and Nestler (2019).

The Random Effects General Cross-Lagged Panel Model (RE-GCLM)

The second kind of next-generation model, called the "general cross-lagged panel model" (GCLM; Zyphur, Allison, et al., 2020) and "random effects GCLM" for generality and coherence with the educational literature (for an explanation of this variant of the GCLM, see Supplementary Material A), can be conceived as a finer-grained version of the RE-CLPM that allows differentiating long-run versus short-run lagged effects, by splitting the within-person variance into two components. More precisely, the lagged effects of an impulse transiting from one construct (e.g., self-beliefs on occasion *T*) to another (e.g., grades on occasion *T+1*), are now allowed to have direct effects via the first construct (effects δ_2 and δ_3 in Figure 4) or indirect effects on a later occasion (e.g., grades on occasion *T+2*) via their persistence in the second construct (effects β'_2 and β'_3 in Figure 4). The fact that short-run and long-run effects are differentiated provides a flexible specification covering a large range of dynamics, including multiple combinations of short-run and long-run effects (small short-run versus large long-run effects, positive short-run versus negative long-run effects, etc.).

—Insert Figure 4 here—

This specification might be appropriate to test the short-run and long-run cross-lagged effects between grades and self-beliefs, in place of the full-forward CLPM that allows for first-order (from *T* to *T+1*) and second-order lagged effects (from *T* to *T+2*; Figure 2, b) and further higher-order effects (from *T* to *T+N*; Arens et al., 2017). Indeed, the full-forward CLPM shows similar limitations as the CLPM (i.e., untenable convergence assumption) and poses additional issues concerning the interpretation of its higher-order lagged effects, which are understood as the direct impact of a distant impulse on a present state. Authors have found this interpretation to be problematic both for methodological reasons (e.g., multicollinearity) and theoretical

reasons (e.g., no rationale exists in the REM theory to assume direct effects of distant impulses on the present state of the system; Ehm et al., 2019). On the contrary, interpreting higher-order effects as the indirect effect of a distant impulse (e.g., on *T-2*) on a present state via its impact on a past state (on *T-1*), as is the case in the GCLM, is compatible with a cumulative view on cross-lagged effects between grades and self-beliefs, in which current states are influenced both by the cross-lagged effect on a prior occasion (on *T-1*) and by the persistence of past cross-lagged effects on the same prior occasion. This specification may therefore identify short-run cross-lagged effects that resorb over time or, conversely, cross-lagged effects that take effects in the long run only. Such diverse dynamics may help to describe the REM in more subtle ways, as a complement to the RE-CLPM.

—Insert Table 2 here—

The Issue of Growth Factors in the Disaggregation of Between- and Within-Person Variance

The above models share the common feature of enabling the estimation of lagged effects while relaxing the convergence assumption. They thus offer a more unequivocal specification of bidirectional effects between two constructs, which is not the case of alternative longitudinal models that confound within- and between-person variances (e.g., CLPM, full-forward model, dual change score model, autoregressive latent trajectory model, fixed effect regression model; Ehm et al., 2019; Mund & Nestler, 2019; Zyphur, Voelkle, et al., 2020). Notwithstanding this common feature, proponents of the RC-CLPM and the GCLM differ in the way they envisage the disaggregation of within- and between-person effects. Indeed, the authors of the GCLM have recommended not including growth factors but keeping a person-specific intercept only (Zyphur, Voelkle, et al., 2020), contrary to what is recommended for the RC-CLPM (Berry & Willoughby,

2017; Curran et al., 2014). According to the former, the variability of a construct over time is hardly decomposable a priori into between-person versus within-person variances. Moreover, they seem to argue that the inclusion of growth factors in the model might overshadow dynamics within lagged effects, by capturing elements of instability in trajectories. Although more research is certainly needed on this issue, one may wonder if the reasons put forward for including a random intercept (i.e., controlling for mean individual differences) apply equally well with regards to the inclusion of growth factors. Indeed, growth factors enable "straightening" up individual trajectories by simulating the existence of a common mean (as in Figure 1, f), thus obtaining a comparable baseline level to ensure convergence in the data. In the present context, this is a necessary specification given that students differ in their growth trajectories of self-concept and achievement (as noted above), and not including growth factors in the model might bias the estimation of lagged-effects, by inducing non-convergence (as in Figure 1, e). A likelihood ratio test can also be envisaged to see whether the inclusion of growth factors improves the fit of the model to the data. For these reasons, the GCLM shown in Figure 4 acknowledges the possibility of including both an intercept and a growth factor (slopes). As before, for terminological and conceptual coherence, Table 2 recognizes random intercept (RI-GLCM) and random curve variants (RC-GLCM) of the GCLM, which all represent "random effects GLCM" (RE-GCLM).

The Added-Value of Next-Generation Cross-Lagged Models in Practice

Although we have argued that many of the lagged-effects methods used to test developmental theories such as the REM are vulnerable to ecological fallacy and are likely to produce spurious findings (due to deviations from the convergence assumption), one might doubt the actual impact of such biases in practice. After all, methods such as the residualized-change

model and the CLPM have been in use for decades in REM research and have always produced consistent findings across different settings and age groups (Arens et al., 2017; Huang, 2011; Marsh & Craven, 2006; Valentine et al., 2004). To better understand the potential limitations of these findings, this section first reviews applications of next-generation cross-lagged models that have revealed methodological issues in standard methods used to test lagged effects. Next, we review similar evidence concerning tests of the REM theory.

Revealing Smushed Effects Due to Non-Convergence

A major added-valued of next-generation models is their capability to reveal dynamics in the data that were previously overshadowed by mean individual differences, thus overcoming the spurious effects of conventional models. In the literature, the latter effects are sometimes called "smushed" effects (Hoffman, 2015, p. 344) to underscore the aggregated nature of the dynamics they embrace, that is, both within- and between-person processes of change. For example, at a given point of time, the variance of grades might be decomposed into a variance determined by individual trajectories (between-person processes of change) and a variance determined by time-specific deviations from individuals' own trajectories (within-person processes of change). When the two kinds of variances exist [i.e., non-convergence setting; Figures 1 b) and e)], a CLPM specification actually forces these dynamics to be aggregated—i.e., smushed—into a single parameter (i.e., the autoregressive effect β_4 in Figure 2, a). Because of this aggregation, smushed effects may mask dramatically different within-person dynamics. For instance, we have described a hypothetical situation in which the CLPM produced spurious significant effects (Type I error) when large mean differences between individuals were not controlled for and covaried between constructs (e.g., correlation between grades and self-beliefs; cf. subsection "Methodological Issues in REM Research"). In other cases, smushed effects may mask the fact

that between-person effects and the within-person effects run in opposite directions (e.g., positive between- and negative within-effects; Hoffman, 2015), thus seriously compromising the research strategy by suggesting positive effects instead of actual negative effects (or conversely), or null effects instead of actual significant effects (Type II error).

Using simulated as well as empirical data from developmental research (e.g., reciprocities between parental spanking and child aggressiveness, parenting style and depressive symptoms, relationship satisfaction and self-esteem) and organizational research (e.g., reciprocities between mean national well-being and mean national income), some studies have indeed shown the existence of multiple kinds of smushed effect by contrasting results from the CLPM and from next-generation models (Berry & Willoughby, 2017; Hamaker et al., 2015; Mund & Nestler, 2019; Zyphur, Allison, et al., 2020). More precisely, their analyses found that cross-lagged effects identified in CLPM may become non-significant or reverse in sign when between-person differences are controlled for using the RI-CLPM, the RC-CLPM or the RI-GCLM; conversely, non-significant cross-lagged effects may become significant. Incidentally, analyses estimating both the RI-CLPM and the RC-CLPM have found similar effects across these models (Mund & Nestler, 2019), suggesting that the controversy surrounding the inclusion of growth factors may be formal rather than substantial. However, the latter analyses omitted controlling for certain covariances between growth factors (i.e., covariances ψ_5 and ψ_6 in Figure 3), thus providing an incomplete specification of the RC-CLPM. More research is therefore needed to understand the differences between the RI-CLPM and the RC-CLPM.

Evidence of Smushed Effects in the REM Research

Two studies have investigated whether the effects identified in the REM research are robust to the biases associated with smushed CLPM effects (Burns et al., 2019; Ehm et al., 2019).

Using a sample of first year psychology students, Burns and his colleagues estimated reciprocal effects between achievement and three measures of self-beliefs (academic competence, self-esteem, self-efficacy) using the CLPM and the RI-CLPM. Their results indicated that, whereas some cross-lagged effects remained largely unchanged across the two models, others disappeared or changed in size, thus indicating the existence of smushed CLPM effects. For example, the cross-lagged effects of achievement on competence was significant in the CLPM, but became not significant in the RI-CLPM. Similarly, the cross-lagged effect of competence on achievement was reduced twofold when moving from the CLPM to the RI-CLPM. Converse changes were found in the cross-lagged effects between self-efficacy and achievement, which appeared small in the CLPM but doubled in size in the RI-CLPM (for more details, see Burns et al., 2019). Unfortunately, this study presents two major issues that render its results quite fragile. First, cross-lagged effects between achievement and the multiple measures of self-beliefs (i.e., academic competence, self-esteem, self-efficacy) were modeled simultaneously, despite the fact that measures of self-beliefs are highly correlated and tend to load on common latent factors (Pietsch et al., 2003). This multicollinearity may pose problems in estimating and interpreting model parameters (Grewal et al., 2004), although the authors provided no information on this issue or on how cross-lagged coefficients behaved when self-beliefs were entered in the models separately. Second, the models were estimated after standardizing the variables to a common mean and standard deviation, a procedure that makes the longitudinal data incomparable across measurement occasions when changes in the sample occur (e.g., due to missing data, as is the case in Burns and colleagues' study) and modifies the function of growth curves (e.g., no evidence for slope when one really exists; Willett, Singer & Martin, 1998). Their standardization procedure might explain why Burns and colleagues found no evidence for

growth (i.e., no slopes) in the constructs under investigation, despite the fact that their study spanned 7 waves of data over a full semester among students experiencing a significant transition (i.e., 1st year of college). Because of these two issues, these findings appear ambiguous and it is unclear whether they constitute solid evidence for smushed effects in REM research.

Using a longitudinal sample of elementary school students spanning four waves of data from grade 1 to grade 4 (N = 2009 students), Ehm and colleagues (2019) also investigated the robustness of the REM theory by using next-generation models. Aligning with the hypothesis of smushed effects, their analyses produced evidence for REM effects when assuming convergence (CLPM), with academic self-concept and achievement (grades, test scores) having positive cross-lagged effects on each other. However, this evidence vanished when the convergence assumption was relaxed using random intercepts for each construct (RI-CLPM), with former significant effects becoming not significant. This suggests that the effects identified in the CLPM specification were spurious and probably confounded with mean individual differences in achievement and self-beliefs. As we found no particular issue with the data treatment, we believe that this study can be taken as solid evidence that the methods usually used to validate the REM theory are not guaranteed.

An Application to the Study of REM Processes Among High School Students

Building on previous sections, we now illustrate the benefits of using next-generation cross-lagged models to study developmental processes. Our literature review argued that the standard model used to test REM processes (CLPM) might be vulnerable to ecological fallacy and specification issues due to its reliance on a strong assumption, that is, that no differences exist between students' growth trajectories in achievement or self-beliefs (convergence assumption). Two alternative kinds of models (RE-CLPM, RE-GCLM) were shown to provide

more adequate modeling strategies, by relaxing this assumption. In this perspective, the following analyses will test the robustness of the REM by comparing results from these alternative strategies (see Table 2 and Figures 2, 3 and 4). Based on our literature review, we expect to validate the REM theory when using the CLPM. Using the RE-CLPM and RE-GCLM will force to disaggregate the between-person and within-person effects that are conflated in the cross-lagged panel model, which often results in cross-lagged effects becoming non-significant or reversing in sign (very seldom do they remain the same). Overall, we therefore expect that the evidence produced by the CLPM to corroborate the REM theory will be spurious in nature (smushed effects) and disappear when the convergence assumption is relaxed.

Method

Participants and Procedure

Participants were 944 French vocational students followed during the first and second years of high school, as part of a longitudinal study on dropout processes. These students were drawn from a larger sample also comprising academic students (n = 1012), but the psychometric analyses for strict longitudinal invariance could only be proven for vocational students and this subsample was therefore selected (convenience sample). Data were collected via self-reported online questionnaires (academic and social background variables, academic self-concept) and administrative databases (grade point average) on a trimestrial basis (semestrial basis for 2nd year grades).

Measures

Grade Point Average. School achievement was measured by the grade point average (GPA) by trimester or semester, as collected by official school records (administrative data). More precisely, GPA refers to the average grade obtained by each student across all disciplines

on a scale of 21 points, 0 being a null score and 20 a perfect score. Student GPA was obtained for each trimester (1st year of high school) and semester (2nd year of high school). Because of internships in the middle of the second year, the latter semesters mostly cover 1st and 3rd trimesters in the year. Overall, this means that GPA was measured for each of the six first trimesters of the study, except for the fifth trimester.

Academic self-concept. Aligning with the REM research, we measured self-beliefs at an intermediate level of specificity, using a scale of academic self-concept (ASC). Our initial measurement scale comprised 5 items from Harter (1988) measured on a 5-point Likert scale (from 1= *Not true at all for me* to 5= *Really true for me*). Psychometric analyses were conducted to explore the factorial structure of the scale and its longitudinal invariance. The analyses showed that negatively-worded items (2 items) and positively-worded items (3 items) could not load on a common latent factor, even after testing multitrait-multimethod strategies (Marsh et al., 2010). However, a 3-item scale was validated (i.e., item 1 = *I have the impression that I am very good in my schoolwork*, item 2 = *I feel as smart as people my age*, item 3 = *I am very good at getting the work done in class*) under strict longitudinal invariance tests (Liu et al., 2017). More precisely, confirmatory factor analyses for the 3-item scale over the 6 waves of measurement showed longitudinal invariance on all psychometric components (i.e., factor loadings, thresholds and variances) and always resulted in satisfactory fit indices (Table S1, Supplementary Material B). This provided evidence that the changes in the observed scores of the scale were driven by changes in the construct of scientific interest (i.e., "alpha" change in academic self-concept), as opposed to changes in the psychometric properties of the scale (i.e., "beta" or "gamma" change in the 3-item scale; Brown, 2015, pp. 221-222). Strict longitudinal invariance is a sufficient condition for estimating models involving growth factors using observed variables (e.g., RC-

CLPM, RC-GCLM; Brown, 2015). Therefore, our measure of academic self-concept is computed by summing up scores from the 3 observed items (for wave 1: Mean = 10.37, Median = 10.00, SD = 2.20, range = [3; 12]). In this scale, higher scores reflect higher levels of academic self-concept. To align with the GPA, we use measurements of academic self-concept on all first 6 trimesters, except for trimester 5.

Covariates. Academic and socio-demographic variables were used to control for confounding effects between REM processes and background variables. More precisely, our models control for *gender* (0 = female adolescent, 1 = male adolescent), *age*, *father SES* (*Low* to *Very High*), *grade retention during primary school* and *during middle school* (1 = yes, 0 = no) and *academic ability* via a national, middle school final exam.

Analytic Strategy

Estimation. Our models account for the fact that the fifth measurement point is missing, by applying equality constraints to the first 4 waves only and by skipping this measurement point in the specification of time (i.e., Time = [0, 1, 2, 3, 5]). We also used Full Information Maximum Likelihood (FIML) to account for missing data (see next paragraph) and a sandwich estimator based on the observed information matrix to obtain results robust to non-normality (i.e., MLR observed; Maydeu-Olivares, 2017). Nested models were tested via a likelihood ratio test, by observing that their deviance did not differ significantly from the deviance of the augmented model according to a scaled chi-square difference test (Satorra & Bentler, 2010). To facilitate the interpretation of parameters while preserving the processes of change intact (Willett et al., 1998), we estimate our models based on unstandardized measures and we report coefficients that were standardized based on the means and standard deviations of endogenous and exogenous variables (endogenous variables only for binary covariates). Models were implemented on R

software (R Core Team, 2016) after adapting and extending existing code (Berry & Willoughby, 2017; Hamaker et al., 2015; Mund & Nestler, 2019; Zyphur, Allison, et al., 2020). As is reported in Table 3, the final models for each REM specification significantly differ from a saturated model at the 5% level (i.e., significant scaled chi-square difference test), which might reflect a misspecification in the models (e.g., student heterogeneity, omitted non-linear parameters; West et al., 2012). This is with the exception of the final RC-GCLM [$\chi^2(88) = 97.9$, $p = .221$]. That being said, all final models show fit indices close to Hu and Bentler's cutoff values (1999), namely CFI ≥ .95, TLI ≥ .95, RMSEA ≤ .06, SRMR ≤ .08, which allows comparing alternative specifications of REM processes.

Model building. Our aim was to obtain a parsimonious specification of REM processes using the six kinds of models presented in Table 2. The general strategy was to obtain a parsimonious model for between-person differences by controlling for latent differences (growth factors) and observed differences (e.g., covariates), before testing equality constraints on the cross-lagged and autoregressive effects, one variable at a time and then simultaneously. More precisely, in step 1, we used likelihood ratio tests and information criteria (Akaike and Bayesian information criteria) to identify bivariate random intercepts (RI-CLPM, RI-GCLM) and latent curves (RC-CLPM, RC-GCLM). Intercepts and slopes were identified for ASC and grades, with covariances between latent factors also contributing to the fit of the model (except for covariances ψ_1, ψ_2 and ψ_5 in Figures 3 and 4). In step 2, we identified significant covariates that predicted first wave measurements (CLPM, full-forward CLPM) or random intercepts (RI-CLPM, RI-GCLM) and latent curves (RC-CLPM, RC-GCLM). As reported in Table 4, all covariates but age and father SES had a significant impact on the latter. In step 3, we identified which lagged effects should be constrained to be equal over time, by observing that this

constraint did not modify the fit of the model significantly (i.e., $p > .05$). All but overall and indirect autoregressive effects of grades (parameters β_4 and β'_4 in Figures 3 and 4) were indeed found to be equal. Finally, in step 4, a parsimonious model was identified that included significant parameters only. In Table 3, we report chi-square difference tests for the unconstrained model (obtained from step 2), the equality constraints model (step 3) and the parsimonious model (step 4). These tests are not significant (e.g., $p > .05$), thus showing that all relevant complexities in the unconstrained model are obtained in the parsimonious models. However, for ease of comparison between models, we focus our discussion of results on the parameters from the "equality constraints" (step 3) models so as to present all lagged effects.

—Insert Table 3 here—

Missing data. Missing data occurred on student background variables (19% for parental SES, 14% for grade retention before high school, 7% for academic ability), our measure of academic self-concept (14% in the first year, 36% in the second year) and on school grades (15% in the first year, 22% in the second year). Application of Little's test (1988) showed that the data was not "missing completely at random", and t-tests and logit analyses showed that missing data was predicted by observed characteristics (i.e., age, gender, grade retention before high school, academic ability) on each measurement point [Mean(NagelkerkeR2)= .597], thus suggesting a possible "missing at random" mechanism. We therefore employed Full Information Maximum Estimation (FIML), a highly-recommended method for handling missing data (Graham, 2012). FIML first computes the maximum likelihood for each individual based on the available data for the variables included in the model, and then combines the results across individuals to compute a full maximum likelihood. To provide more robust estimates, the variables associated with missingness were also included in the models as determinants of the observed variables for the

first measurement point or as determinants of random effects. By using this strategy as well as the information contained across multiple measurement occasions, our models are estimated on the basis of 100% of our sample and should be robust to biases induced by missing data.

Results

Identification of Differences Between Individual Trajectories of Academic Self-Concept and Grades

To relax the convergence assumption, we first identified individual differences in the development of academic self-concept and grades. As can be seen from the random effects CLPM (RI-CLPM, RC-CLPM) in Table 4 and from the random effects GCLM (RI-GCLM, RC-GCLM) in Table 5, the intercept and slope of each construct appear to be significant, which means that high school students in our sample differ both in terms of the level and the shape of their development trajectories in grades and academic self-concept. Both kinds of random effects should therefore be included to ensure convergence in the data. We also find that random effects covary between constructs. More precisely, students experiencing high (low) initial grades also report high (low) initial academic self-concept (i.e., positive covariance ψ_4 between intercepts). Also, students initially high (low) in academic self-concept tend to experience more negative (less negative) declines in grades over time (i.e., negative covariance ψ_6 between the academic self-concept intercept and the negative slope of grades), which might be evidence that the initial dispersion in grades regresses towards the mean over time. Finally, the positive covariance ψ_3 between the negative and positive slopes of grades and academic self-concept indicates that students who experience increases in grades over time also experience decreases in ASC, and conversely (compensatory covariance). Although the latter observation might be taken as evidence against REM processes—in the sense that gains in one construct are associated with

losses on the other construct (cf. "Discussion")-—, the causal framework privileged in the literature (i.e., Granger causality) focuses on cross-lagged effects, to which we turn next.

Identification of REM Processes As Short-Run Lagged Effects

In this section, we present results on REM processes as identified under alternative specifications of short-run, cross-lagged effects from one measurement to the next. Following our literature review (Figures 2 and 3), we use the symbols β_2 and β_3 to represent, respectively, the standardized cross-lagged effect of academic self-concept on grade point average and its reciprocal effect.

As hypothesized, specifying REM processes within the CLPM specification replicates the results from the literature (Table 4). On the one hand, prior academic self-concept has a positive effect on subsequent grade point average both during the first year ($\beta_{2,T1-T4}$ = [.039; .047], p = .043) and second year of high school ($\beta_{2,T4-T6}$ = .047, p = .090). These effects might be considered too small to be meaningful (i.e., β_2< .05; Keith, 2015), but they still reflect a significant positive influence on grade point average. Likewise, prior grade point average has a positive effect on subsequent academic self-concept during the first year ($\beta_{3,T1-T4}$ = [.214; .226], p < .001) and second year of high school ($\beta_{3,T4-T6}$ = .245, p < .001). These estimates align with the meta-analysis of Valentine and colleagues (2004) who found a stronger effect size for grades on academic self-concept (β_3 = .16) than for academic self-concept on grades (β_2 = .08), and they also correspond with the estimates reported in one of the most recent studies on REM processes, which studied reciprocities between math self-concept and math grades (β_2 = [.045; .049], β_3 = [.134; .166]; Arens et al., 2017). In summary, our data gives support to the REM theory when using the CLPM.

—Insert Table 4 here—

However, we argued that the CLPM confuses processes of change situated within the person and those situated between people. As hypothesized on the basis of methodological studies (Berry & Willoughby, 2017; Hamaker et al., 2015; Mund & Nestler, 2019), disaggregating both processes via the random-effects CLPM changes the cross-lagged effects of interest and qualifies the REM theory (Table 4). More precisely, when mean individual differences are controlled for via a random intercept (RI-CLPM), the effect of prior academic self-concept on subsequent grade point average appears to be meaningful in size (i.e., β_2>.05; Keith, 2015), but this effect is in fact not significant in the first year of high school ($\beta_{2,T1-T4}$ = [.060; .079], p = .286), albeit it is significant during the second year ($\beta_{2,T4-T6}$ = .111, p = .033). Moreover, evidence is found for the reciprocal effect of grades on self-concept during the first year ($\beta_{3,T1-T4}$ = [.184; .202], p = .021), as well as during the second year of high school ($\beta_{3,T4-T6}$ = .188, p = .035). Taken together, these results suggest that the evidence previously provided by the CLPM was partly distorted due to the conflation or smushing of within- and between-person processes of change, although the REM finds full validation during the second year of high school in the RI-CLPM. Alternatively, when mean individual differences are controlled for by growth factors in addition to a random intercept (i.e., slopes for ASC and grades in the RC-CLPM), all cross-lagged effects become non-significant and do not contribute to model fit anymore [$\chi^2_{diff}(8)$ = 23.2, p = .917; reduced model not shown]. To the extent that the growth factors and their covariances all contribute to model fit [$\chi^2_{diff}(2)$ = 13.4, p = .001; reduced model not shown] and are significant in the parsimonious model (i.e., without non-significant lagged effect), these results suggest that the effects previously identified in the RI-CLPM were due to a lack of control in mean individual differences related to the shape of trajectories between individuals or, in other words, to an insufficient relaxation of the

convergence assumption. As is apparent from Table 3, the best model for short-run lagged effects appears to be the RC-CLPM (vs. CLPM, RI-CLPM), both in terms of fit indices (i.e., CFI, TLI, RMSEA, SRMR) and in terms of deviance from a saturated model [lowest $\chi^2(92)$ = 126.1, p = .011].

Identification of REM Processes As Short-Run Versus Long-Run Lagged Effects

We now identify REM processes using specifications for both short-run and long-run cross-lagged effects as represented, respectively, by the beta prime and delta coefficients in Figure 4. We also report results for the full-forward model that also includes short-run and long-run effects, although we discussed the theoretical and methodological problems associated with this model (see subsection "The Random Effects General Cross-Lagged Model").

Using the full-forward CLPM, we replicate results from the literature that find no significant REM effects in second-order relationships (i.e., from *T* to *T+2;* Arens et al., 2017; Ehm et al., 2019), and their inclusion cancels out first-order relationships (see Open Data). As discussed by the latter authors, this is likely due to the fact that higher-order relationships may have no incidence over and above first-order relationships (from *T* to *T+1*), for the very reason that the latter is itself influenced by the former, thus leading to redundancies in model specifications. On the contrary, the random effects GCLM (RI-GCLM, RC-GCLM) presented in Table 5 overcome these limitations by assuming direct and indirect effects that are not redundant in their influences. More precisely, the RI-GLCM suggests that, although self-perceptions have a positive effect on grades in the short-run both during the first year ($\delta_{2,T1-T4}$ = [.259; .309], p < .001) and second year of high school ($\delta_{2,T4-T6}$ = .558, p < .001), these effects are compensated in the long-run by negative indirect effects on grades of a similar intensity ($\beta'_{2,T1-T4}$ = [-.378; -.341], p < .001; $\beta'_{2,T4-T6}$ = -.567, p < .001). No such effects are identified

for the cross-lagged effects of grades on self-perceptions, which all prove to be non-significant. Adding controls for slopes, as in the RC-GCLM, does not change these observations once we cancel out non-significant lagged effects (cf. parsimonious RC-GCLM, Table 4). However, the short-run and long-run effects are considerably smaller in this specification, which suggests that part of the compensatory effects in the RI-GLCM were in fact driven by between-person variance (i.e., individual differences in slopes). Because the controls for slopes provide a better fit to the model [$\chi^2_{diff}(5) = 254.0$, $p < .001$], we are inclined to prefer the RC-GCLM (vs. RI-GCLM) as a better modeling strategy to relax the convergence assumption. In this configuration, the overall cross-lagged effects of self-perceptions, which are computed by adding up their long-run and short-run cross-lagged effects (Zyphur, Allison, et al., 2020), are non-significant ($\beta'_{2,T1-T4} + \delta_{2,T1-T4} = -.007$, $p = .830$; $\beta'_{2,T4-T6} + \delta_{2,T4-T6} = .014$, $p = .366$). The latter observation aligns with results from the RC-CLPM that aggregated these effects. Importantly, the fact that the final RC-GCLM presents a non-significant difference from a saturated model [$\chi^2\ (88) = 97.9$, $p = .221$] suggests that the parameters in this model are capable of reproducing the processes that generated the observed data in all its complexities, at least from a demographic point of view (i.e., developmental system at the population level).

—Insert Table 5 here—

Overall, these analyses illustrate a situation where effects identified in the CLPM are spurious in nature because of their conflation of within- and between-person processes of change, as revealed by next-generation cross-lagged models. The comparison of next-generation models also gives support to the view that, in order to relax the convergence assumption effectively, researchers should test whether the inclusion of growth factors (e.g., linear, quadratic or cubic random effects) contributes to ameliorate the model over and above random intercepts. Indeed,

not including relevant growth factors may lead to identify spurious cross-lagged effects (as in our application of the RI-CLPM) or to overestimate their size (as in our application of the RI-GCLM). According to our best model specification (RC-GCLM), we can conclude from this kind of analyses that gains in grades have no effect on gains in self-beliefs, whereas gains in self-beliefs have a positive effect on gains in grades in the short-run, but this positive effect is offset by their negative effect in the long-run. The latter conflict between short-run versus long-run effects results in the observation that the overall effect of self-beliefs on grades and the reverse effect are both non-significant.

Discussion

Summary of Themes

The present article offered a review of modeling methods aimed at identifying reciprocal causal effects in non-experimental designs, namely standard cross-lagged effects models (i.e., cross-lagged panel model, residualized change model), versus next-generation cross-lagged effects models (i.e., the random effects CLPM and GCLM; Curran et al., 2014; Hamaker et al., 2015; Zyphur, Allison, et al., 2020). Although the review focused on the theory of reciprocal effects between self-beliefs and school achievement (REM theory), the review is also relevant to other fields of education that might have used standard cross-lagged effects models.

In essence, the review built around the statistical assumption of convergence, which states that individual differences in growth trajectories are null so that, by implication, processes of change within individuals (i.e., changes relative to oneself) reduce to processes of change between individuals (i.e., changes relative to others). We outlined the reasons why standard cross-lagged effects models provide valid evidence on reciprocities insofar as the convergence assumption is met. When this is not the case, standard methods provide unreliable evidence that

conflates processes of change within- versus between- individuals, making the research strategy vulnerable to ecological fallacy (i.e., within-person processes of change are inferred from evidence partly based on between-person processes of change) and to statistical artifacts (i.e., the regression coefficients are byproducts of the conflation, also known as "smushed" effects). Next-generation models were then presented that relax the convergence assumption by controlling for individual differences in growth trajectories, thus providing a more robust identification strategy.

Confronting standard versus next-generation cross-lagged models reveals insufficiencies in the research conducted on developmental processes. Many studies have now shown that well-established effects in the literature (i.e., replicated across many studies via standard lagged effects models and confirmed by meta-analyses) may in fact disappear or reverse in sign, once the convergence assumption is met via the RE-CLPM or the RE-GCLM. On the contrary, non-significant effects previously rejected by empirical studies may turn out to be, in fact, significant (Berry & Willoughby, 2017; Hamaker et al., 2015; Mund & Nestler, 2019; Zyphur, Allison, et al., 2020). As the present analyses and others have shown (Ehm et al., 2019), the evidence on the REM theory also appears to be vulnerable to the smushed effects generated by standard cross-lagged models. More precisely, results from CLPM analyses replicate the well-established finding that school achievement and self-beliefs have mutually positive effects over time, yet these effects disappear when specifying REM processes using the RE-CLPM or the RE-GCLM.

Theoretical and Methodological Implications for Research on School Motivation

The developmental processes claimed by the REM theory are found, more generally, in major theories of school motivation, including theories of self-efficacy, self-regulation, self-worth, self-affirmation, self-determination, etc. (cf. section "Literature Review on the Reciprocal Effects Model"; Valentine et al., 2004). In this perspective, it is plausible that the biases

identified for the REM theory in the present article concern these other school motivation theories as well. Although complementary research is certainly needed to further understand the nature of these biases, this preliminary conclusion casts doubts on the state of knowledge in the field and invites us to reconsider the way developmental processes are described and investigated.

First, educational psychologists might benefit from integrating the distinction between intra-individual and inter-individual processes of change, especially on a conceptual level (Nesselroade, 1991). As the present study and other studies have shown (Berry & Willoughby, 2017; Hamaker et al., 2015; Mund & Nestler, 2019; Zyphur, Allison, et al., 2020), the two kinds of processes seldom overlap and might even run in opposite directions. Yet, with the notable exception of researchers working on growth trajectories (see hereafter "Limitations of Next-Generation Cross-Lagged Models"), their distinction has not been sufficiently recognized in the literature. Indeed, although theories describe how changes in motivation or other states within the student cause changes in behavior within the same student (intra-individual change), their empirical evidence is based on standard cross-lagged models that confound causes of change situated between and within individuals (smushed inter- and intra- individual change; Christenson et al., 2012; Wentzel & Wigfield, 2009). This confusion is certainly exacerbated by the fact that developmental theories are depicted in schematic representations that closely resemble cross-lagged models, thus supporting the illusion that the intra-individual processes described by the theory correspond to the processes instantiated in such models. However, as the random effects CLPM specifications show (Curran et al., 2014; Hamaker et al., 2015), the changes identified by cross-lagged models are more adequately described as *deviations from* initial levels or expected growth trajectories. This brings an additional layer of complexity for which schematic representations need to be developed in order to depict—and therefore

distinguish—the intra-individual and inter-individual processes of change simultaneously. More generally, this means that efforts should be made to account, on a theoretical level, for the methodological insights brought forth by multilevel models (including next-generation cross-lagged models) concerning the distinction of these two kinds of processes.

Second, to the extent that previously identified lagged effects may become non-significant or reverse in sign when the convergence assumption is met (e.g., using RE-CLPM or RE-GLCM specifications), we might consider recognizing the limitations of previous evidence used to validate school motivation theories. Three *hypothetical* situations can be envisaged to understand these limitations. The first situation concerns processes of change for which no individual differences in growth trajectories exist (i.e., convergence), or for which the within-person variance prevails over the between-person variance. In this case, previous evidence based on standard cross-lagged methods should be taken at face value to describe within-person processes of change. The second situation concerns processes of change for which individual differences exist but are constant over time (i.e., non-convergence). In this case, if the between-person variance prevails over within-person variance, then previous evidence will prove to be spurious in nature, because the identified lagged effects will actually be driven by unvarying differences between individuals (i.e., no developmental gains or losses in the constructs under study; see, for an example concerning REM processes, Ehm et al., 2019). Finally, the third situation concerns processes of change for which individual differences exist and vary over time, meaning they differ as a function of a random intercept and additional random growth factors. Here, if the between-person variance prevails over the within-person variance, then previous evidence needs to be re-examined by using alternative strategies to identify reciprocities in terms of between-person processes of change (cf. next subsection).

In this perspective, and if the data requirements allow it (i.e., at least four waves of data), researchers might envisage testing the robustness of previous findings via next-generation cross-lagged models, by testing for the existence of differences between individual trajectories of growth. Until then, the uncertainty surrounding the evidence based on standard lagged effects models may be too important to give credit to replication studies or to meta-analyses based on standard lagged effects models. This insight is both damaging and exciting for the field, for although it might be seen as a recess in the dynamics of knowledge accumulation, it also opens new areas of research with unforeseen, beneficial implications for educational interventions.

Limitations of Next-Generation Cross-Lagged Panel Models

The epistemological and methodological critiques discussed in this article have been demonstrated before and apply particularly well to the REM research, as the present analyses have shown. Notwithstanding their relevance, we also recognize that next-generation models (RE-CLPM, RE-GCLM) have their own limitations. First, these models adhere to a "cartesian-newtonian" or "efficient" view on causality (i.e., the Granger causality) that recognizes causes as unfolding mechanically between isolated entities (i.e., effects between independent variables in a static system), as opposed to holistically (i.e., co-dependent variables in a dynamical system; Molenaar, 2001; Overton, 2015). In next-generation models, this reductionist view on causality implies identifying lagged effects that, for statistical reasons explained in this article (the convergence assumption), ultimately apply to intraindividual variability, that is, to *deviations from* individual trajectories of change. We may wonder whether motivational theories really consider dynamics associated with these deviations or, alternatively, with the individual trajectories themselves. In the latter case, novel cross-lagged models that apply to intraindividual change are needed to enable tests of the REM within this reductionist approach.

Contrary to the cross-lagged effects privileged by REM researchers, researchers investigating reciprocities between self-beliefs and achievement might also be interested in the growth factors of these constructs and in their covariances, which does not align with a Granger-like, reductionist approach. In this case, alternative views on causality may be needed. For example, the framework of "formal" causality recognizes that the structural constraints in a system (e.g., initial levels of school achievement and academic self-concept) limit the kinds of processes allowed within the system (e.g., changes in school achievement and academic self-concept), in a way that tends to stabilize the system in the form of homeostatic functions (Molenaar, 2001; Overton, 2015). Within this framework, covariances between trajectories of growth in self-beliefs and achievement might be considered as a evidence on REM processes understood as formal causalities (vs. efficient causalities). Although a meta-analysis is needed on this issue, the present study as well as a few studies (Aunola et al., 2002; Fraine et al., 2007) found the covariance between slopes of achievement and academic self-concept to be negative or null, which does not support the notion that gains in one construct lead to gains in the other.

Second, our applications of next-generation models may be faulted for not investigating REM processes from a multi-group perspective, which might have otherwise revealed that the REM processes are actually observed among specific subgroups of students defined by observed characteristics (e.g., gender, age grade; multi-group modeling) or unobserved characteristics (mixture modeling). We decided not to investigate this question because our goal was to illustrate the different model specifications and their contribution to the analysis of processes of change, while at the same time adopting the conventional approach on REM processes for which no developmental subgroups are expected. Future research might nevertheless investigate this issue further by using multi-group applications of next-general models and, possibly, find

subgroups of students for whom an REM framework applies, thus pointing to specific educational interventions.

Third, next-generation models may be faulted for being vulnerable to another kind of ecological fallacy (i.e., not directly related to the convergence assumption), which is to assume that the parameters identified for intra-individual change in the population of students provide evidence for developmental processes at the level of each individual student, that is, the assumption of homogeneity of the ensemble (Molenaar et al., 2003). Future research might therefore be conducted to develop school motivation theories that describe the nature of the gap between demographic and idiographic processes of change, and to model them concomitantly by assessing which dynamics are generalizable to the population of students, and which are not. Models based on dynamic structural equations (DSEM) provide an appropriate framework to this end (Asparouhov et al., 2017, 2018). In particular, these models allow testing whether lagged effects are individual-specific or invariant in the population, thus offering a means to model ensemble homogeneity in processes of change. In contrast to next-generation cross-lagged model, the within-person component of the DSEM relates to individual-specific lagged effects (vs. population-aggregated lagged effects), and its between-person component to the growth trajectories of the lagged effects (vs. growth trajectories of the construct under investigation). Because the lagged effects are based on time-series data and may include an arbitrary number of effects from past states of the developmental system, the DSEM offers a means to model co-dependencies or homeostatic functions that undergird stationarity or bifurcations in the system (Molenaar, 2001), thus offering the conceptual tools to a holistic or "relation-developmental-systems" view on causality (Overton, 2015, p. 9). However, the DSEM requires intensive time-series data that is often unavailable in educational research (≥ 10 waves of data), and it also

implies additional complexities such as validating the longitudinal invariance of scales at an individual-specific level, whereas current scales for motivational constructs have been validated at a population-specific level. Considering these limitations, the REM theory might still be open for validation over and above the statistical artifacts and ecological fallacies discussed before.

Conclusion

The present article contributes to the field of educational psychology by showing that popular modeling strategies currently in use (i.e., autoregressive models such as the residualized change model or the cross-lagged panel model) provide fragile evidence for the investigation of developmental processes, notably by producing spurious results (statistical artifacts) or information not directly relevant to the level of analysis usually investigated, that is, processes of change within the individual (ecological fallacy). Moreover, we illustrated how these shortcomings could be overcome by using alternative specifications that provide more robust findings on the developmental processes portrayed in theory. Using more appropriate modeling strategies, we found no support for the claim that students who experience increases in school achievement (i.e., grade point average) also experience, later on, increases in self-beliefs (i.e., academic self-concept), and we found little support for the reciprocal effect. The theory according to which self-beliefs and school achievement mutually reinforce each other over time (i.e., the REM theory) appears instead to be based on evidence vulnerable to statistical artifacts and ecological fallacy. Because this theory undergirds most theories of school motivation, the critique probably extends to these as well. In this sense, the present article might encourage educational psychologists to question the state of knowledge in the field by revising their modeling strategies (e.g., cross-lagged models, growth curve models, dynamic models) and their privileged causal frameworks (e.g., efficient versus holistic causality), thus propelling new

research avenues.

References

Arens, A. K., Marsh, H. W., Pekrun, R., Lichtenfeld, S., Murayama, K., & vom Hofe, R. (2017). Math self-concept, grades, and achievement test scores: Long-term reciprocal effects across five waves and three achievement tracks. *Journal of Educational Psychology*, *109*(5), 621. https://doi.org/10.1037/edu0000163

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2017). Dynamic latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(2), 257–269. https://doi.org/10.1080/10705511.2016.1253479

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 359–388. https://doi.org/10.1080/10705511.2017.1406803

Aunola, K., Leskinen, E., Onatsu-Arvilommi, T., & Nurmi, J.-E. (2002). Three methods for studying developmental change: A case of reading skills and self-concept. *British Journal of Educational Psychology*, *72*(3), 343–364. https://doi.org/10.1348/000709902320634447

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.

Berry, D., & Willoughby, M. T. (2017). On the practical interpretability of cross-lagged panel models: Rethinking a developmental workhorse. *Child Development*, *88*(4), 1186–1206. https://doi.org/10.1111/cdev.12660

Bollen, K. A. (1989). *Structural equations with latent variables* (Vol. 210). John Wiley & Sons.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*

(Vol. 467). John Wiley & Sons.

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, *15*(1), 1–40. https://doi.org/10.1023/A:1021302408382

Brandt, N. D., Mike, A., & Jackson, J. J. (2019). Do school-related experiences impact personality? Selection and socialization effects of impulse control. *Developmental Psychology*, *55*(12), 2561. https://doi.org/10.1037/dev0000817

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

Burns, R. A., Crisp, D. A., & Burns, R. B. (2019). Re-examining the reciprocal effects model of self-concept, self-efficacy, and academic achievement in a comparison of the Cross-Lagged Panel and Random-Intercept Cross-Lagged Panel frameworks. *British Journal of Educational Psychology*. https://doi.org/10.1111/bjep.12265

Cai, Y., King, R. B., Law, W., & McInerney, D. M. (2019). Which comes first? Modeling the relationships among future goals, metacognitive strategies and academic achievement using multilevel cross-lagged SEM. *Learning and Individual Differences*, *74*, 101750. https://doi.org/10.1016/j.lindif.2019.06.004

Christenson, S. L., Reschly, A. L., & Wylie, C. (Eds.). (2012). *Handbook of research on student engagement*. Springer Science & Business Media. DOI 10.1007/978-1-4614-2018-7

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, *62*(1), 583–619. https://doi.org/10.1146/annurev.psych.093008.100356

Curran, P. J., Howard, A. L., Bainter, S. A., Lane, S. T., & McGinley, J. S. (2014). The separation of between-person and within-person components of individual change over

time: A latent curve model with structured residuals. *Journal of Consulting and Clinical Psychology*, *82*(5), 879–894. https://doi.org/10.1037/a0035297

Curran, P. J., & Wirth, R. J. (2004). Interindividual differences in intraindividual variation: Balancing internal and external validity. *Measurement: Interdisciplinary Research and Perspectives*, *2*(4), 219.

Ehm, J.-H., Hasselhorn, M., & Schmiedek, F. (2019). Analyzing the developmental relation of academic self-concept and achievement in elementary school children: Alternative models point to different results. *Developmental Psychology*, *55*(11), 2336–2351. https://doi.org/10.1037/dev0000796

Engels, M. C., Pakarinen, E., Lerkkanen, M.-K., & Verschueren, K. (2019). Students' academic and emotional adjustment during the transition from primary to secondary school: A cross-lagged study. *Journal of School Psychology*, *76*, 140–158. https://doi.org/10.1016/j.jsp.2019.07.012

Fraine, B. D., Damme, J. V., & Onghena, P. (2007). A longitudinal analysis of gender differences in academic self-concept and language achievement: A multivariate multilevel latent growth approach. *Contemporary Educational Psychology*, *32*(1), 132–150. https://doi.org/10.1016/j.cedpsych.2006.10.005

Garon-Carrier, G., Boivin, M., Guay, F., Kovas, Y., Dionne, G., Lemelin, J.-P., Séguin, J. R., Vitaro, F., & Tremblay, R. E. (2016). Intrinsic motivation and achievement in mathematics in elementary school: A longitudinal investigation of their association. *Child Development*, *87*(1), 165–175. https://doi.org/10.1111/cdev.12458

Graham, J. W. (2012). *Missing data: Analysis and design*. Springer Science & Business Media.

Grewal, R., Cote, J. A., & Baumgartner, H. (2004). Multicollinearity and measurement error in

structural equation models: Implications for theory testing. *Marketing Science*, *23*(4), 519–529. https://doi.org/10.1287/mksc.1040.0070

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116. https://doi.org/10.1037/a0038889

Harter, S. (1988). *Manual for the self-perception profile for adolescents*. University of Denver.

Harter, S. (1999). *The construction of the self: A developmental perspective*. Guilford Press.

Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. Routledge.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945–960.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, *49*(5), 505–528. https://doi.org/10.1016/j.jsp.2011.07.001

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5–86. https://doi.org/10.1257/jel.47.1.5

Keith, T. Z. (2015). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling* (2nd ed.). Routledge.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202.

https://doi.org/10.1080/01621459.1988.10478722

Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development*, *31*(4), 357–365. https://doi.org/10.1177/0165025407077757

Liu, Y., Millsap, R. E., West, S. G., Tein, J., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods*, *22*(3), 486–506. https://doi.org/10.1037/met0000075

Marsh, H. W. (1986). Global self-esteem: Its relation to specific facets of self-concept and their importance. *Journal of Personality and Social Psychology*, *51*(6), 1224–1236. https://doi.org/10.1037/0022-3514.51.6.1224

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, *1*(2), 133–163. https://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, *81*(1), 59–77. https://doi.org/10.1348/000709910X503501

Marsh, H. W., & O'Mara, A. (2008). Reciprocal effects between academic self-concept, self-esteem, achievement, and attainment over seven adolescent years: Unidimensional and multidimensional perspectives of self-concept. *Personality and Social Psychology Bulletin*, *34*(4), 542–552. https://doi.org/10.1177/0146167207312313

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking

jingle-jangle fallacies. *Journal of Educational Psychology*, *111*(2), 331–353. https://doi.org/10.1037/edu0000281

Marsh, H. W., Scalas, L. F., & Nagengast, B. (2010). Longitudinal tests of competing factor structures for the Rosenberg Self-Esteem Scale: Traits, ephemeral artifacts, and stable response styles. *Psychological Assessment*, *22*(2), 366. https://doi.org/10.1037/a0019225

Maydeu-Olivares, A. (2017). Maximum likelihood estimation of structural equation models for continuous data: Standard errors and goodness of fit. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(3), 383–394. https://doi.org/10.1080/10705511.2016.1269606

Molenaar, P. C. M. (2001). Systems modeling. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 15423–15428). Pergamon Press.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Molenaar, P. C. M. (2008). On the implications of the classical ergodic theorems: Analysis of developmental processes has to focus on intra-individual variation. *Developmental Psychobiology*, *50*(1), 60–69. https://doi.org/10.1002/dev.20262

Molenaar, P. C. M., Huizenga, H. M., & Nesselroade, J. R. (2003). The relationship between the structure of interindividual and intraindividual variability: A theoretical and empirical vindication of developmental systems theory. In U. M. Staudinger & U. Lindenberger (Eds.), *Understanding human development* (pp. 339–360). Springer. https://doi.org/10.1007/978-1-4615-0357-6_15

Mund, M., & Nestler, S. (2019). Beyond the cross-lagged panel model: Next-generation statistical tools for analyzing interdependencies across the life course. *Advances in Life Course Research*, *41*, 100249. https://doi.org/10.1016/j.alcr.2018.10.002

Nagy, G., Watt, H. M. G., Eccles, J. S., Trautwein, U., Lüdtke, O., & Baumert, J. (2010). The development of students' mathematics self-concept in relation to gender: Different countries, different trajectories? *Journal of Research on Adolescence*, *20*(2), 482–506. https://doi.org/10.1111/j.1532-7795.2010.00644.x

Nesselroade, J. R. (1991). The warp and the woof of the developmental fabric. In R. M. Downs, L. S. Liben, & D. S. Palermo (Eds.), *Visions of aesthetics, the environment, and development: The legacy of Joachim F. Wohlwill* (pp. 213–240). Lawrence Erlbaum Associates, Inc.

Overton, W. F. (2015). Processes, relations, and relational-developmental-systems. In W. F. Overton & P. C. M. Molenaar (Eds.), *Handbook of child psychology and developmental science* (Vol. 1, pp. 9–62). John Wiley & Sons, Inc. http://onlinelibrary.wiley.com/doi/10.1002/9781118963418.childpsy102/full

Pajares, F., & Schunk, D. H. (2001). Self-beliefs and school success: Self-efficacy, self-concept, and school achievement. In R. J. Riding & S. G. Rayner (Eds.), *Self perception* (Vol. 2, pp. 239–266). Ablex Publishing.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*, 422. https://doi.org/10.3389/fpsyg.2017.00422

Parker, P. D., Marsh, H. W., Morin, A. J. S., Seaton, M., & Zanden, B. V. (2015). If one goes up the other must come down: Examining ipsative relationships between math and English self-concept trajectories across high school. *British Journal of Educational Psychology*,

85(2), 172–191. https://doi.org/10.1111/bjep.12050

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 259–282). Springer Science & Business Media.

Pfost, M., Hattie, J., Dörfler, T., & Artelt, C. (2014). Individual differences in reading development: A review of 25 years of empirical research on Matthew effects in reading. *Review of Educational Research*, *84*(2), 203–244. https://doi.org/10.3102/0034654313509492

Pietsch, J., Walker, R., & Chapman, E. (2003). The relationship among self-concept, self-efficacy, and performance in mathematics during secondary school. *Journal of Educational Psychology*, *95*(3), 589. https://doi.org/10.1037/0022-0663.95.3.589

Pitzer, J., & Skinner, E. A. (2016). Predictors of changes in students' motivational resilience over the school year: The roles of teacher support, self-appraisals, and emotional reactivity. *International Journal of Behavioral Development*, *41*(1), 15–29. https://doi.org/10.1177/0165025416642051

Quin, D. (2017). Longitudinal and contextual associations between teacher–student relationships and student engagement: A systematic review. *Review of Educational Research*, *87*(2), 345–387. https://doi.org/10.3102/0034654316669434

R Core Team. (2016). *R: A language and environment for statistical computing* (3.6) [Computer software]. R Foundations for Statistical Computing. https://www.R-project.org

Reeve, J. (2012). A self-determination theory perspective on student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 149–172). Springer Science & Business Media.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*(3), 351–357.

Ruzek, E. A., & Schenke, K. (2019). The tenuous link between classroom perceptions and motivation: A within-person longitudinal study. *Journal of Educational Psychology*, *111*(5), 903–917. https://doi.org/10.1037/edu0000323

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*(2), 243–248. https://doi.org/10.1007/s11336-009-9135-y

Selig, J. P., & Little, T. D. (2012). Autoregressive and cross-lagged panel analysis for longitudinal data. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 265–278). The Guilford Press.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford university press.

Skinner, E. A. (1996). A guide to constructs of control. *Journal of Personality and Social Psychology*, *71*(3), 549–570. https://doi.org/10.1037/0022-3514.71.3.549

Sutter-Brandenberger, C. C., Hagenauer, G., & Hascher, T. (2018). Students' self-determined motivation and negative emotions in mathematics in lower secondary education—Investigating reciprocal relations. *Contemporary Educational Psychology*, *55*, 166–175. https://doi.org/10.1016/j.cedpsych.2018.10.002

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

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, *39*(2), 111–133. https://doi.org/10.1207/s15326985ep3902_3

Vasconcellos, D., Parker, P. D., Hilland, T., Cinelli, R., Owen, K. B., Kapsal, N., Lee, J., Antczak, D., Ntoumanis, N., Ryan, R. M., & Lonsdale, C. (2019). Self-determination theory applied to physical education: A systematic review and meta-analysis. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000420

Weidinger, A. F., Steinmayr, R., & Spinath, B. (2017). Math grades and intrinsic motivation in elementary school: A longitudinal investigation of their association. *British Journal of Educational Psychology*, *87*(2), 187–204. https://doi.org/10.1111/bjep.12143

Wentzel, K. R., & Wigfield, A. (2009). *Handbook of motivation at school*. Routledge.

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (Vol. 1, pp. 209–231). The Guilford Press.

Willett, J. B., Singer, J. D., & Martin, N. C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations. *Development and Psychopathology*, *10*(2), 395–426. https://doi.org/10.1017/S0954579498001667

Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z., Hamaker, E. L., Shamsollahi, A., Pierides, D. C., Koval, P., & Diener, E. (2020). From data to causes I: Building a general cross-lagged panel model (GCLM). *Organizational Research Methods*, *23*(4), 651–687. https://doi.org/10.1177/1094428119847278

Zyphur, M. J., Voelkle, M. C., Tay, L., Allison, P. D., Preacher, K. J., Zhang, Z., Hamaker, E. L., Shamsollahi, A., Pierides, D. C., Koval, P., & Diener, E. (2020). From data to causes II: Comparing approaches to panel data analysis. *Organizational Research Methods*, *23*(4), 688–716. https://doi.org/10.1177/1094428119847280

Table 1

Examples of REM Effects in School Motivation Theories

Motivational Theory	Effect of self-belief on academic performance (self-enhancement hypothesis)	Effect of academic performance on self-belief (skill-development hypothesis)
Self-efficacy (Pajares & Schunk, 2001)	Feeling confident about one's ability to perform a given task induces more engagement during task performance (increased perseverance, cognitive and metacognitive strategies) which, in turn, increases the quality of performance on the task	Performing well on a task (being skillful) induces mastery experiences which, in turn, increases feelings of self-efficacy
Self-determination (Pitzer & Skinner, 2016; Reeve, 2012)	Feeling competent at school induces high-quality school engagement (in terms of behaviors, emotions, cognitions) which, in turn, increases academic achievement	High-quality academic achievement elicits feelings of competency as well as teacher and parental support which, in turn, increases feelings of competency
Self-regulation (Panadero, 2017)	Feeling competent at school induces a more adaptive preparative phase (e.g., setting mastery goals, task representation and planning) which, in turn, increases task performance	Performing well during a task facilitates positive self-evaluations of one's competencies
Achievement emotion processes (Pekrun, 2006; Pekrun & Linnenbrink-Garcia, 2012)	Feeling in control in school activities deemed important facilitates positive achievement emotions (e.g., joy, hope, pride) which, in turn, increases academic achievement	High-quality achievement elicits feelings of control as well as positive feedbacks and consequences on the learning environment which, in turn, increases feelings of control

Note. Processes and constructs are described in broad terms so as to include multiple variants of each motivational theory (for details, please see the references). Effects are understood as causal influences of one construct onto another.

Table 2

Standard and Next-Generation Cross-Lagged Effects Models

	Controls used to relax the convergence assumption		
Kinds of lagged effects	No controls	Random Intercept	Random Curve
Short-Run Lagged Effects (*T* to *T+1*)	CLPM	RI-CLPM	RC-CLPM ("LCM-SR")
Short-Run and Long-Run Lagged Effects (*T* to *T+1, T* to *T+2*)	Full-forward CLPM	RI-GLCM ("GCLM")	RC-GCLM

Note. CLPM= cross-lagged panel model; GCLM= general cross-lagged model; RI= random intercept; RC= random curve; "GCLM"= general cross-lagged model (Zyphur, Allison, et al., 2020); LCM-SR= latent curve model with structured residuals (Curran et al., 2014). Standard models in the first column assume convergence between individual trajectories (CLPM), whereas next-generation models in the second and third columns relax the convergence assumption by controlling for differences in developmental trajectories using random effects at the individual level (i.e., random intercepts and growth factors for latent curves).

Table 3

Fit Indices for REM Processes Between Academic Self-Concept and Grade Point Average

Model	χ^2	df	χ^2_{diff}	Δdf	RMSEA	SRMR	CFI	TLI
CLPM								
1	416.8***	77	—	—	0.074	0.071	0.905	0.871
2	**414.4*** **	**83**	**3.2**	**6**	**0.071**	**0.072**	**0.906**	**0.881**
3	416.1***	85	2.6	2	0.070	0.072	0.906	0.884
RI-CLPM								
1	372.8***	85	—	—	0.062	0.057	0.927	0.909
2	**365.7*** **	**91**	**2.3**	**6**	**0.059**	**0.059**	**0.928**	**0.917**
3	361.1***	94	2.3	3	0.058	0.059	0.928	0.920
RC-CLPM								
1	121.9***	73	—	—	0.028	0.038	0.987	0.982
2	**122.0** **	**79**	**2.2**	**6**	**0.025**	**0.038**	**0.988**	**0.985**
3	126.1*	92	7.8	13	0.022	0.041	0.99	0.989
FF-CLPM								
1	183.4***	65	—	—	0.046	0.042	0.968	0.949
2	189.9***	74	10.3	9	0.044	0.047	0.968	0.954
3	194.1***	78	4.3	4	0.043	0.047	0.968	0.957
RI-GCLM								
1	116.4***	69	—	—	0.028	0.042	0.987	0.981
2	**132.0*** **	**81**	**16.0**	**12**	**0.027**	**0.046**	**0.986**	**0.982**
3	152.0***	94	20.2†	13	0.028	0.049	0.984	0.982
RC-GCLM								
1	63.0	57	—	—	0.011	0.026	0.998	0.997
2	**87.7†**	**71**	**23.5†**	**14**	**0.017**	**0.042**	**0.995**	**0.993**
3	**97.9**	**88**	**12.0**	**17**	**0.012**	**0.040**	**0.997**	**0.997**

Note. 1= unconstrained model; 2= model with equality constraints; 3= model with significant parameters only. Models in bold characters correspond to those reported in Table 4 and Table 5.

†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 4

Alternative Models of REM Processes Specifying Short-Run Lagged Effects

Parameters	CLPM		RI-CLPM		RC-CLPM	
Autoregressive Effects (β_1, β_4)	*ASC*	*GPA*	*ASC*	*GPA*	*ASC*	*GPA*
T1-T2	.354***	.784***	-.024	.270*	-.036	.049
T2-T3	.373***	.846***	-.025	.502***	-.040	.331***
T3-T4	.363***	.832***	-.024	.412**	-.039	0082
T4-T6	.361***	.846***	.080	.541***	-.066	-.132
Cross-Lagged Effects (β_2, β_3)	*ASC => GPA*	*GPA => ASC*	*ASC => GPA*	*GPA => ASC*	*ASC => GPA*	*GPA => ASC*
T1-T2	.039*	.214***	.060	.202*	.023	.084
T2-T3	.041*	.226***	.062	.193*	.027	.090
T3-T4	.047*	.219***	.079	.184*	.039	.084
T4-T6	.047†	.245***	.111*	.188*	.186	-.014
Random Effects and Covariances			*ASC*	*GPA*	*ASC*	*GPA*
Intercept			5.245***	3.674***	4.802***	3.677***
Slope					0.587†	-0.911*
ψ_1, ψ_2						
ψ_3, ψ_4				.481***	.478	.558***
ψ_5, ψ_6						-.098
Covariates	*ASC (T1)*	*GPA (T1)*	*ASC*	*GPA*	*ASC*	*GPA*
Gender (1 = male)	.175*	-.259***		-.398***	.240*	-.271***
					-.579**	-.365***
GR primary school		-.162*		-.276***		-.166*
						-.378**
GR middle school	.230**	.246**				.204*
						-.322**
Academic Ability	.204***	.498***	.275***	.582***	.278***	.544***
						.243***

Note. ASC= Academic Self-Concept; GPA= Grade Point Average; GR= Grade Retention. For the RI-CLPM and the RC-CLPM, the effects of covariates relate to the intercept (1st line) and slope (2nd line), respectively. Models include covariances between measurements (cf. Table S3, SM). Names of coefficients and covariances correspond to those reported in Figure 2 and Figure 3. Raw estimates and standard errors reported in Table S4, SM. †$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 5

Alternative Models of REM Processes Specifying Short-Run and Long-Run Lagged Effects

Parameters	RI-GCLM		RC-GCLM		RC-GCLM (parsimonious)	
LR Autoregressive Effects (β_1', β_4')	*ASC*	*GPA*	*ASC*	*GPA*	*ASC*	*GPA*
T1-T2	.018	.270***	-.023	.084		
T2-T3	.018	.282***	-.024	.087		
T3-T4	.017	.290***	-.023	.089		
T4-T6	-.187	.403***	-.266†	.135		
SR Autoregressive Effects (δ_1, δ_4)	*ASC*	*GPA*	*ASC*	*GPA*	*ASC*	*GPA*
T1-T2	-.023	-.017	-.034	-.025		
T2-T3	-.025	.020	-.037	.078†		.123***
T3-T4	-.025	-.078	-.036	-.037		
T4-T6	.260†	-.021	.147	-.084		
LR Cross-Lagged Effects (β_2', β_3')	*ASC => GPA*	*GPA => ASC*	*ASC => GPA*	*GPA => ASC*	*ASC => GPA*	*GPA => ASC*
T1-T2	-.341***	-.034	-.132	.008	-.036***	
T2-T3	-.355***	-.035	-.137	.009	-.037***	
T3-T4	-.378***	-.033	-.142	.008	-.038***	
T4-T6	-.567***	.113	-.218	.195	-.052***	
SR Cross-Lagged Effects (δ_2, δ_3)	*ASC => GPA*	*GPA => ASC*	*ASC => GPA*	*GPA => ASC*	*ASC => GPA*	*GPA => ASC*
T1-T2	.259***	-.024	.111	.046	.029†	
T2-T3	.283***	-.023	.123	.047	.032†	
T3-T4	.309***	-.020	.127	.042	.033†	
T4-T6	.558***	.088	.215†	-.084	.066*	
Random Effects and Covariances	*ASC*	*GPA*	*ASC*	*GPA*	*ASC*	*GPA*
Intercept	5.425***	4.056***	4.843***	3.720***	4.884***	3.643***
Slope			0.412	-0.408	0.569†	-0.639†
ψ_1, ψ_2						
ψ_3, ψ_4		.745***	.568***	.510†	.703***	.604***
ψ_5, ψ_6				.034		-.135†

Note. ASC= Academic Self-Concept; GPA= Grade Point Average. Models comprise covariances between measurements and the same covariates as in Table 4 (cf. Table S5, SM). Names of coefficients and covariances correspond to those reported in Figure 4. Raw estimates and standard errors reported in Table S6, SM. †p < .10. *p < .05. **p < .01. ***p < .001.

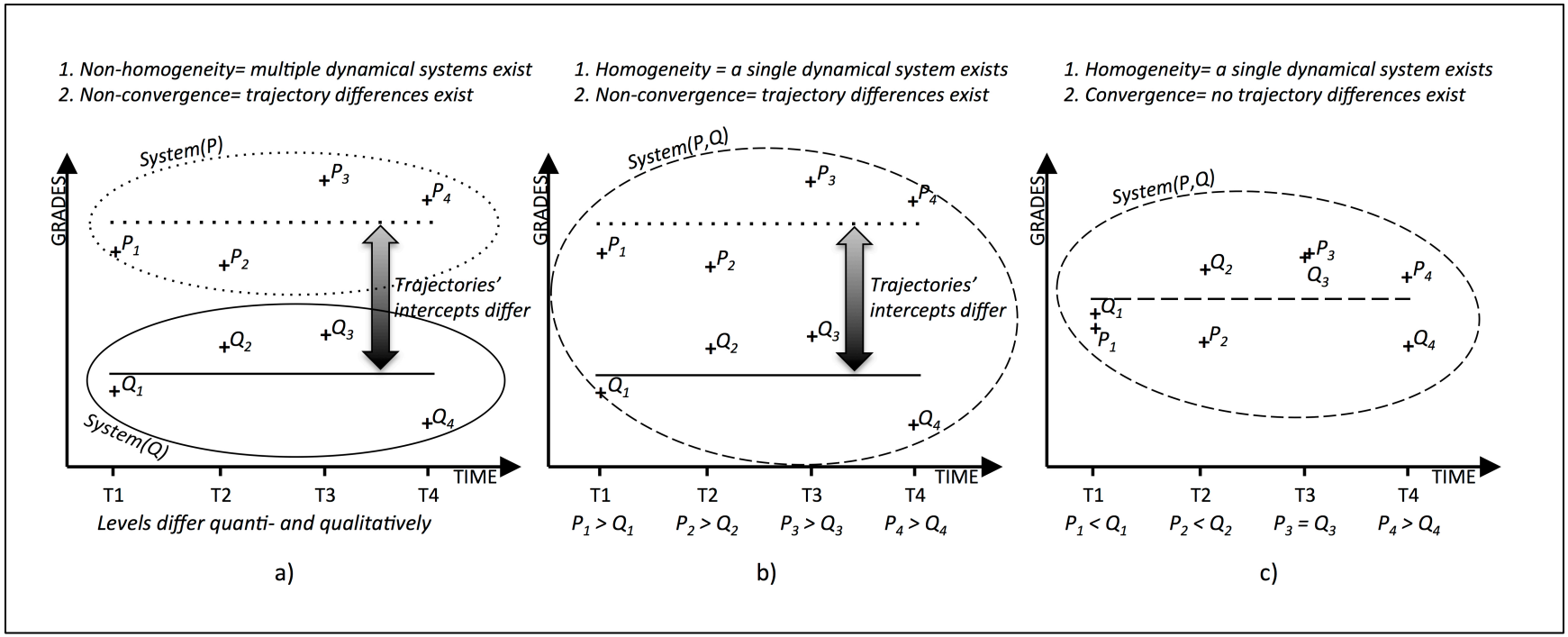


Figure 1. Illustration of Homogeneity and Convergence Assumptions in Processes of Change

Note. Data points represent grades obtained by individuals P and Q, on four measurement occasions (T1, T2, T3 and T4). Horizontal lines represent individual trajectories [with varying intercepts in a) and b), plus varying slopes in d) and e)], whereas ellipses represent the dynamical systems driving the processes of change. For each series of graphs [a) to e), and e) to f)], graphs differ in the degree to which they assume homogeneity (a single dynamical system applies for all individuals) and convergence (no trajectory differences exist between individuals) in the processes of change in grades. We assume that the processes of change are

stationary, so that situations of homogeneity imply ergodicity.

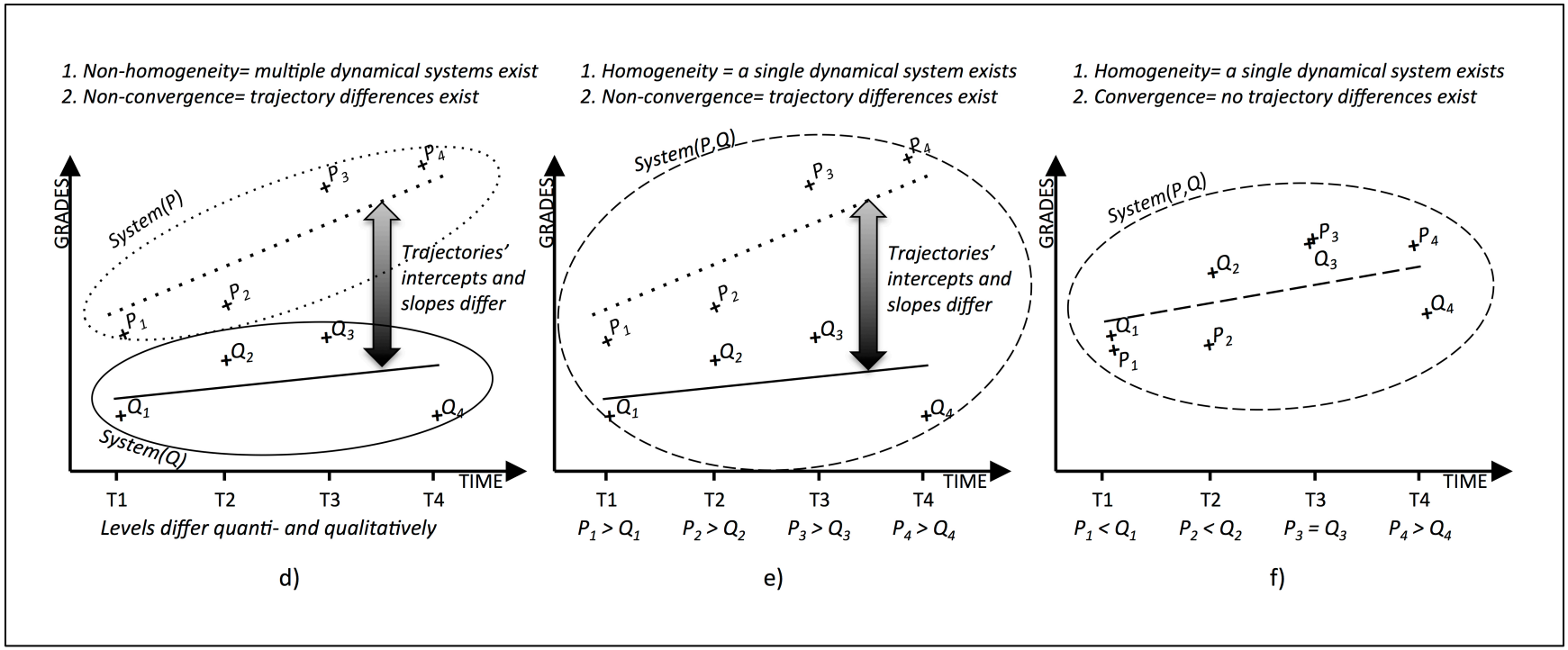
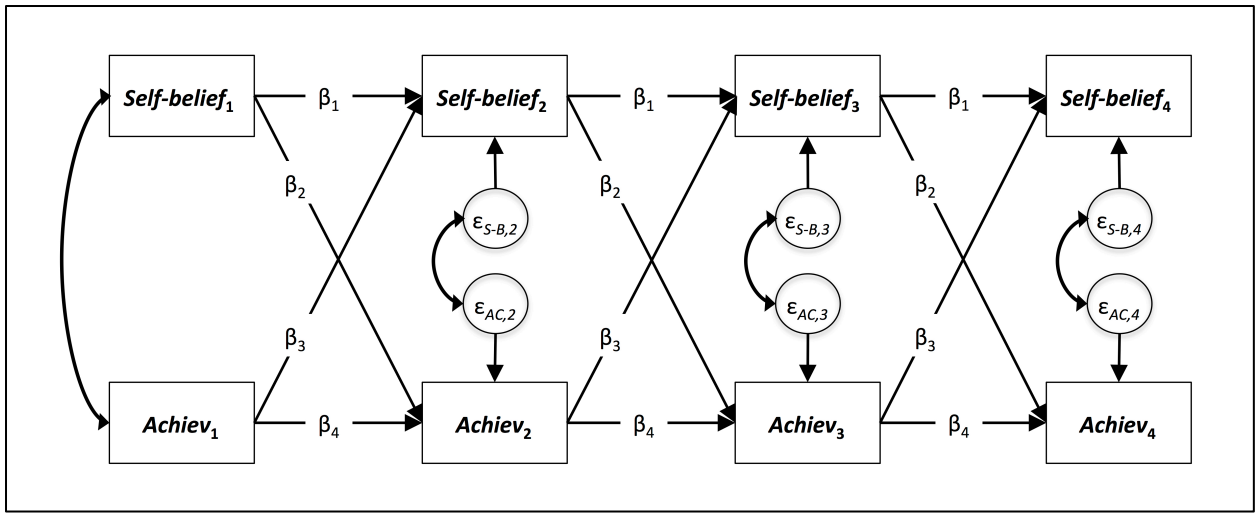
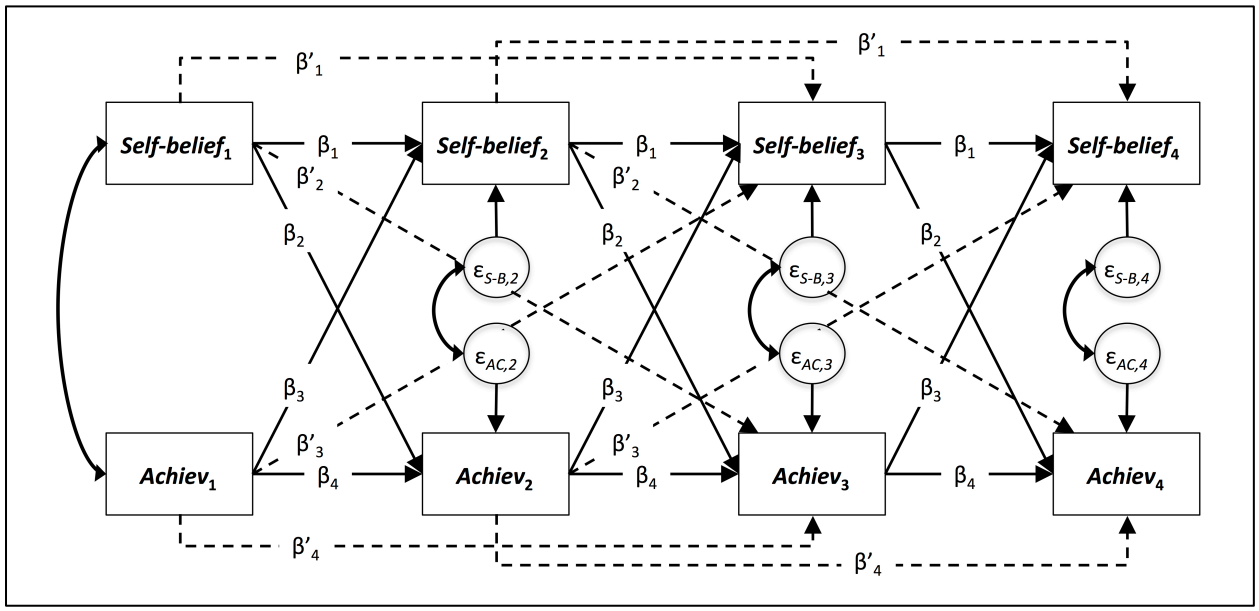


Figure 1 (Continued).



a)



b)

Figure 2. REM Processes As Specified in the Cross-Lagged Panel Model.

Note. Variants of the model include (a) first-order lagged effects (*T* to *T+1*) and (b) first-order lagged effects (*T* to *T+1*) and second-order lagged effects (*T* to *T+2*). Models are schematic and omit some parameters (mean and variance structures; for a complete specification, see the Open Data attached to the article).

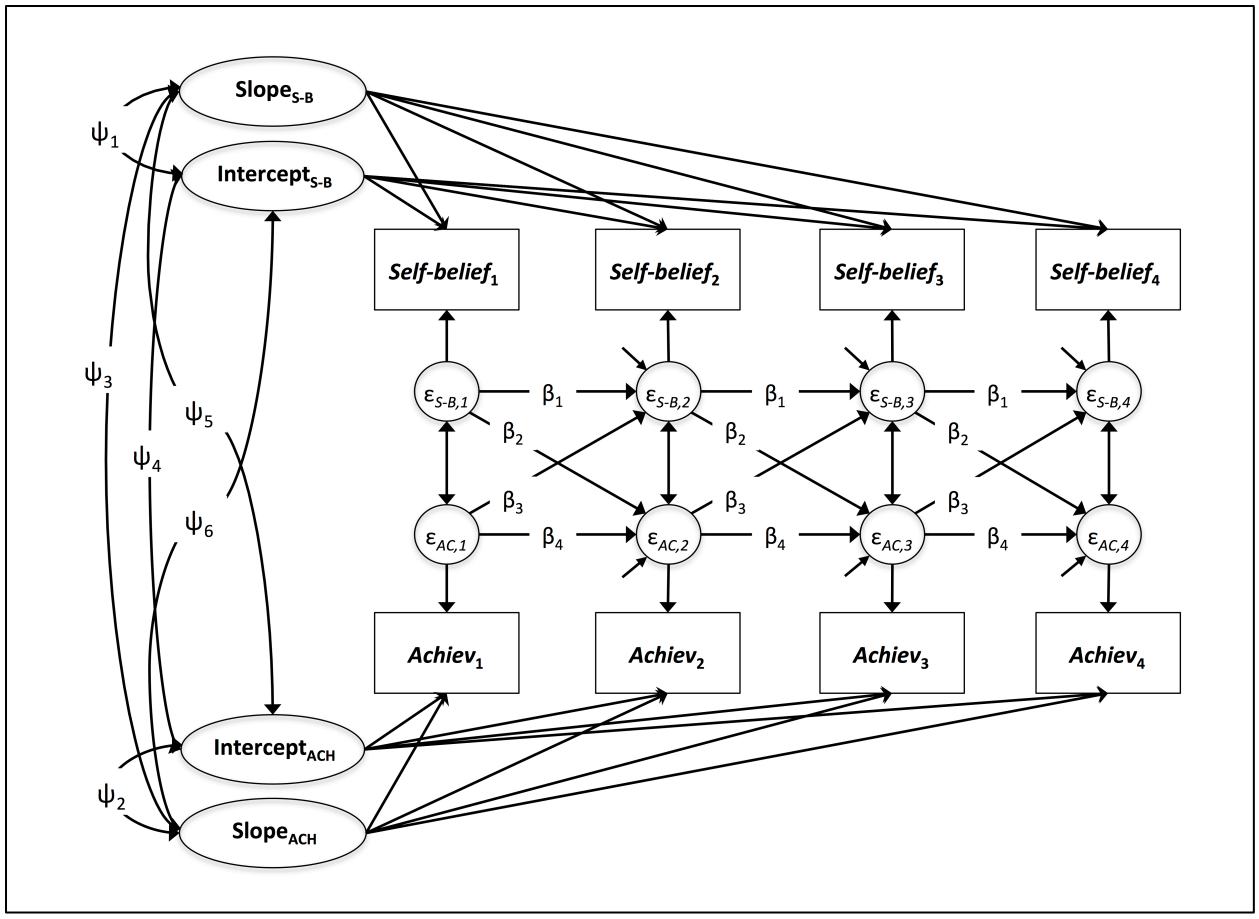


Figure 3. REM Processes As Specified in the Random Curve Cross-Lagged Panel Model

Note. Intercept effects are fixed to 1, whereas slope effects are incremental (e.g., effects equal to 0, 1, 2 and 3 across measurements 1 through 4, assuming equal spacing between measurements and centering on the first measurement occasion). Residual intercepts are fixed to zero to identify growth parameters. Other model parameters are omitted for simplicity (mean and variance structures; for a complete specification, see the Open Data attached to the article).

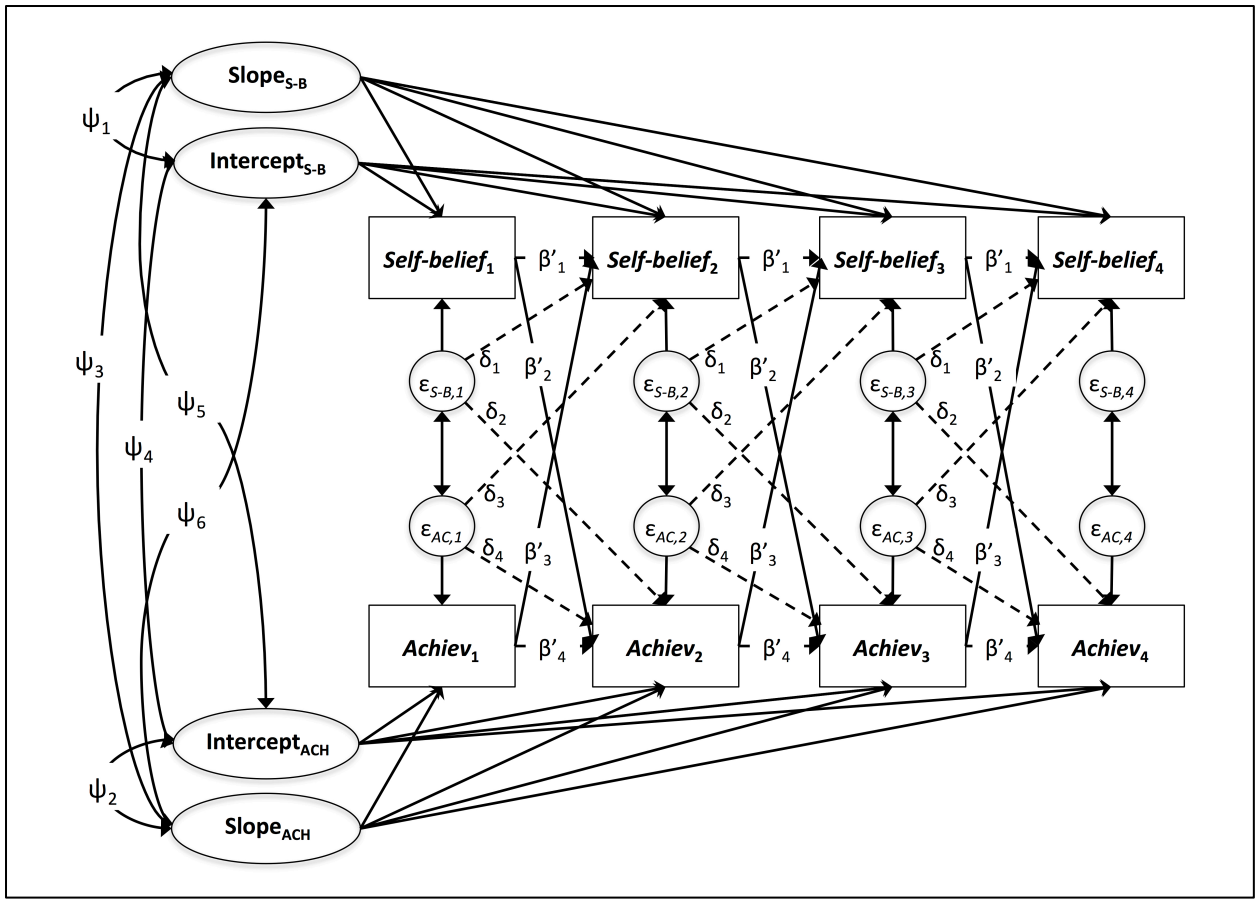


Figure 4. REM Processes As Specified in the Random Curve General Cross-Lagged Panel Model

Note. Intercept effects are fixed to 1, whereas slope effects are incremental (e.g., effects equal to 0, 1, 2 and 3 across measurements 1 through 4, assuming equal spacing between measurements and centering on the first measurement occasion). Residual intercepts are fixed to zero to identify growth parameters. Other model parameters are omitted for simplicity (mean and variance structures; for a complete specification, see the Open Data attached to the article).