School Leaders' Self-Efficacy and Job Satisfaction Over Nine Annual Waves:

A Substantive-Methodological Synergy Juxtaposing Competing Models of Directional Ordering

Herbert W. Marsh[a], Oliver Lüdtke[b], Reinhard Pekrun[a,c], Philip D. Parker[a], Kou Murayama[d], Jiesi Guo[a] ,

Geetanjali Basarkod[a], Theresa Dicke[a], James N. Donald[e] , Alexandre J.S. Morin[f]

[a]Australian Catholic University, [b] Leibniz Institute for Science and Mathematics Education (IPN),[c] University of Essex, [d] University of Tübingen, [e]The University of Sydney, [f] Concordia University

Contemporary Educational Psychology, in press

Date of acceptance: March 09, 2023

Citation for this article:

Marsh, H. W., Lüdtke, O., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Basarkod, G., Dicke, T., Donald, J. N., Morin, A. J. S. (in press). School leaders' self-efficacy and job satisfaction over nine annual waves: A substantive-methodological synergy juxtaposing competing models of directional ordering. *Contemporary Educational Psychology.*

School Leaders' Self-Efficacy and Job Satisfaction Over Nine Annual Waves: A Substantive-Methodological Synergy Juxtaposing Competing Models of Directional Ordering

Highlights

1. The school principal's job is becoming increasingly demanding and complex, but their well-being is under-studied.
2. Job satisfaction and self-efficacy are key constructs for school principal's well-being.
3. Research assumes self-efficacy leads to job satisfaction but is untested.
4. School-leader job satisfaction and self-efficacy are reciprocally related over time.
5. Methodologically we critique and juxtapose competing statistical models of reciprocal effects.
6. Our study is a substantive-methodological synergy of strong data, methodology, theory & implications.

Keywords:
self-efficacy;
job satisfaction';
school principal health and well-being;
cross-lagged panel models of reciprocal effects;
measurement invariance and stationarity;
within- and between-person perspectives

Abstract

The school principal's job is increasingly demanding and complex, but school-principal well-being is understudied. Self-efficacy and job satisfaction are critical constructs for studying school principals' well-being, and self-efficacy is a core predictor of job satisfaction. Cross-sectional research typically assumes a unidirectional ordering; self-efficacy predicts (and leads to) job satisfaction, not the reverse. However, this unidirectional ordering is inconsistent with theoretical models positing a bidirectional (reciprocal) ordering. Furthermore, the assumption is largely untested with appropriate longitudinal data and statistical models. We evaluated the directional ordering of job satisfaction and self-efficacy for a large (N = 5663), nationally representative, longitudinal (nine annual waves) sample of Australian school leaders. Job satisfaction and self-efficacy were moderately correlated within waves and over time. Consistently with theoretical models and a priori predictions, the two constructs were reciprocally related over time; prior measures of each had small statistically positive effects on subsequent measures of the other, with no evidence of directional predominance of one over the other. Support for reciprocal effects was remarkably consistent across competing cross-lag-panel models, multiple tests of the consistency of effects over time (measurement invariance and stationarity), control for covariates, and the addition of lag-2 paths. Methodologically, we critique competing models that estimate cross-lagged effects and evaluate directional ordering from withinand between-person perspectives. We demonstrate the value of both approaches in achieving a robust framework for assessing longitudinal panel models.. Our substantive-methodological synergy has important substantive implications for theory, policy, and practice—showing that school-leader job satisfaction and selfefficacy are mutually reinforcing.

Keywords:
self-efficacy;
job satisfaction';
school principal health and well-being;
cross-lagged panel models of reciprocal effects;
measurement invariance and stationarity;
within- and between-person perspectives

School Leaders' Self-Efficacy and Job Satisfaction Over Nine Annual Waves: A Substantive-Methodological Synergy Juxtaposing Competing Models of Directional Ordering

The role of the school principal is becoming increasingly demanding with responsibility for personnel, school economy, facilities, instructional leadership, and staff well-being, as well as student achievement, development, and well-being (Grissom et al., 2019; Hallinger, 2011; Leithwood & Louis, 2012; Rilely, 2020; Skaalvik, 2020a, b). Increasingly, schools face a crisis as aging school leaders seek early retirement because of these high demands and suitable candidates reluctant to take their place (Riley, 2020; Tran et al., 2018). A school principals' self-efficacy is vital to their success and predictive of their job satisfaction (Federici & Skaalvik, 2011; 2012; also see Darmody& Smyth, 2016; Leithwood et al.,2008; 2011; Liu et al., 2021; Skaalvik 2020a,b; Skaalvik & Skaalvik, 2017). Beyond education settings, job satisfaction is one of the most widely studied constructs in applied psychology and management science, and self-efficacy is a core construct in predicting job satisfaction (e.g., Judge et al. 2020; Kasalak & Dagyar, 2020).

Based on largely cross-sectional research, the relation between job satisfaction and self-efficacy is typically characterized by a unidirectional ordering in which self-efficacy predicts (and leads to) job satisfaction, and not the reverse (see Federic & Skaalvik, 2011, 2012; Wu and Griffin, 2012). However, this unidirectional characterization is inconsistent with Bandura's (1977; 1986; 1989; 1997) theory of reciprocal determinism which states that self-efficacy and outcomes are reciprocally related over time. Furthermore, related theoretical models used in applied research are also consistent with the Bandura assumption of reciprocal effects (e.g., Bakker et al., 2014, 2017; Fredrickson, 2001; 2008; Hobfoll, 1989). Thus, Skaalvik (2020b; also see Federici & Skaalvik, 2012) noted the need for longitudinal research to evaluate the possibility of reciprocal relations involving school leaders' self-efficacy. Hence, although there is strong evidence that these constructs are moderately correlated, there is no compelling evidence that the directional ordering flows only from self-efficacy to job satisfaction.

The traditional cross-lag-panel model (CLPM) has been the dominant model used to test the directional ordering of associations between variables relevant to theoretical predictions. More recently, there has been considerable interest in a variation of the CLPM with random intercepts (RI-CLPM). The RI-CLPM provides a within-person perspective posited to provide a more robust basis for inferences about directional ordering (e.g., Hamaker et al., 2015; Mulder & Hamaker, 2021). The underlying conceptual and statistical issues posed by CLPMs and RI-CLPMs are very different and address fundamentally different questions. The relative appropriateness of between-person (CLPM) and within-person (RI-CLPM) and their interpretations are currently hot substantive and methodological topics. Here we argue that both the CLPM and RI-CLPM perspectives are relevant and that it is helpful to contrast results based on these alternative perspectives.

Our study is a substantive-methodological synergy (Marsh & Hau, 2007; also see Hoffmans et al., 2021). We evaluate the job satisfaction and self-efficacy relation for a large (N = 5,663), longitudinal (nine annual waves), nationally representative sample of Australian school leaders. Substantively, we test for reciprocal effects relating job satisfaction and self-efficacy over this extended time period. Methodologically, our study is one of the first large-scale applications juxtaposing fully-latent versions of CLPMs and RI-CLPMs that overcomes many limitations of manifest models. Our literature review begins with brief overviews of the job satisfaction and self-efficacy constructs and the relations between the two. We then introduce the statistical models that are the basis of our research. Our substantive-methodological synergy focuses on substantively important issues that have implications for theory, research, policy, and practice. Methodologically, we extend existing methodological frameworks that broadly apply to educational psychology research. Finally, bringing together these substantive and methodological foci, we demonstrate how to conceptualize and test the directional ordering of job satisfaction and self-efficacy.

Substantive Focus on Job Satisfaction and Self-Efficacy

Job Satisfaction

Judge et al. (2020) note that job satisfaction is the most widely studied job attitude and possibly the most widely studied topic in the history of applied psychology (Judge et al., 2017). Ock (2020) emphasized that organizations want to satisfy their employees (Spector, 1997). In addition, job satisfaction is related to important outcomes, including turnover (Schleicher et al., 2011; Tett & Meyer, 1993), absenteeism (Hackett, 1989), and counterproductive work behavior (Berry et al., 2012).

Mincu (2015; also see Judge et al., 2017) defines job satisfaction as “a positive frame of mind that is reflected by the employee’s opinion regarding work or the climate of his workplace;” an affective orientation toward one's job (Federici & Skaalvik, 2012). However, it is possible to operationalize “overall” job satisfaction in different ways: as a relatively unidimensional global factor or a summative score (perhaps weighted by importance) that incorporates specific components (e.g., facets such as pay, supervision, and work condition; Marsh & Scalas, 2018; also see Federic & Skaalvik, 2012; Judge et al. 2017; Macdonald & Maclntyre, 1997). Macdonald and Maclntyre reviewed studies suggesting that facets are merely components of an overarching, more global factor. They concluded that the general approach better represents the overall satisfaction. Relatedly, Federici & Skaalvik (2012) noted that the facet approach to measuring school principals' job satisfaction might overlook critical facets and the differential importance assigned to the various factors. Our study's measure of job satisfaction is a hybrid of these two approaches. It includes both a global component (how satisfied are you with your job as a whole, everything taken into consideration) and specific components (work prospects, working conditions, use of your abilities). In this sense, we treat job satisfaction as a unidimensional latent construct that includes both global and specific facets.

Psychological constructs typically vary along a state-trait continuum and usually contain aspects of both (e.g., Cattell, 1966; Dicke et al., 2021; Dormann et al., 2006; Kenny & Zautra, 2001). Even relatively state-like constructs (e.g., mood) have some trait-like components. Conversely, even relatively fixed traits (e.g., personality traits and IQ) are somewhat malleable and can evolve over time. We conceive job satisfaction as a moderately enduring trait closer to the middle of the continuum (e.g., Dormann et al., 2006), but not as fixed as, for example, personality traits or intelligence. This is justified, for instance, by the substantial test-retest correlation of job satisfaction responses from one year to the next. Indeed, as we emphasize, this is central to the distinction between the CLPM and the RI-CLPM.

Self-Efficacy

Self-efficacy is defined as “beliefs in one’s capabilities to mobilize the motivation, cognitive resources, and courses of action needed to meet given situational demands” (Wood & Bandura, 1989, p. 408). When setbacks occur, high self-efficacy individuals recover more quickly (Schwarzer & Hallum, 2008). Following Bandura's (1977, 1997) social learning theory, self-efficacy is broadly defined as confidence in executing courses of action in managing applicable in a wide array of situations. Self-efficacy expectations are a key predictor of future behavior and have considerable explanatory power with a wide range of behaviors. Self-efficacy integrates performance experience, vicarious experiences, imaginal experience, verbal persuasion, and physiological and emotional states (Bandura, 1977, 1997; Maddux, 2009; Maddux et al., 2012).

Self-efficacy measures have historically focused on the performance of specific tasks—domain- or situationally-specific efficacy beliefs. However, subsequent extensions have suggested general self-efficacy concerning a more broadly based set of situations. It is, of course, relevant to test the generalizability of predictions based on Bandura's (1977) conceptual model to more generalized measures of self-efficacy. This is because generalized self-efficacy is built on the generalized, repeated experiences of task-specific characteristics over different situations. In this sense, the theoretical rationale is the same for more generalized and task-specific measures. Indeed, Bandura (1977) referred explicitly to "generalized expectations of personal efficacy " (p. 200) and noted, "Efficacy expectations also differ in generality. Some experiences create circumscribed mastery expectations. Others instill a more generalized sense of efficacy that extends well beyond the specific treatment situation" (p. 194).

Consistently with Bandura's reciprocal causation principle in relation to generalized self-efficacy, Sherer et al. (p. 670) suggested that: "Individuals with high self-efficacy expectations are more likely to attempt new behaviors and to persist in them, and in turn are more likely to meet with successes, thereby increasing their self-efficacy expectations." Based on this extension of Bandura's social cognitive theory (SCT), generalized job satisfaction and self-efficacy are predicted to be reciprocally related. Furthermore, the specificity-matching principle (e.g., Brunswick, 1952; Swan et al., 2007) suggests that a predictor's breadth and generality should match the outcome it is designed to predict. Thus, it is appropriate to use highly specific self-efficacy items in relation to highly task-specific outcomes. However, using more generalized self-efficacy measures for more broadly generalized outcomes such as job satisfaction is more appropriate. Relatedly, Pajares (1996) warned that "microscopically operationalized" measures of self-efficacy lose practical utility, even as predictive power increases. Hence, for broad, generic constructs like job satisfaction, a generalized measure of self-efficacy is likely to be more predictive than an overly specific task measure of self-efficacy. Schwarzer and Hallum (2008) found that task-specific self-efficacy measures and their Generalized Self-efficacy scale (like what we used in the present investigation) were highly correlated and similarly related to job stress and burnout. In a longitudinal analysis, they found that task-specific and generalized measures of self-efficacy measures, rather than mapping unto distinct constructs actually defined a single self-efficacy latent construct.

In educational research, Federici and Skaalvik (2012; also see Hannah et al., 2008) lamented the lack of attention given to school principals' self-efficacy and problems with available instruments. A common conceptualization of principal self-efficacy is the self-perception of a principal's own "capabilities to structure a particular course of action in order to produce desired outcomes in the school he or she leads"

(Tschannen-Moran & Gareis 2004, p. 573). However, as Skaalvick (2020a) noted, school leaders have increasing responsibility for all that happens within the school. Thus, this overarching definition of principal self-efficacy does not guide how to measure principal self-efficacy. As with job satisfaction, self-efficacy can be measured with a relatively global measure like the Schwarz and Hallum (2008) generalized measure of self-efficacy or self-efficacy concerning narrowly defined, domain-specific facets.

In the facet approach to principal self-efficacy, instruments relied on measuring a small number of domain-specific facets; for example, instructional leadership, management, moral leadership (Tschannen-Moran & Gareis 2004; 2009); instructional program, school mission, learning climate (Hallinger & Murphy, 1985); planning, curriculum/teaching, staff management, budgeting, managing parents, and school environment (Dimmock & Hattie, 1996); clarifying educational goals and expectations, promoting teacher learning, resourcing, planning and evaluating teaching and curriculum, and supportive environment (Hallinger, 2010); develop goals, guide teachers, learning environment, motivate teachers, develop a collective culture (Skaalvik, 2020a). However, given the very broad and generic nature of the school principal's job, there is no consensus on what facets should be included or even if the same facets are appropriate in different educational jurisdictions. Furthermore, in research based on their facet model of principal self-efficacy, Skaalvik (2020a, 2020b, Federici & Skaalvick, 2012) treated the specific facets as indicators of a higher-order factor of global self-efficacy that was then related to other constructs (e.g., global job satisfaction). Nevertheless, as Federici (2012; also see Skaalvik, 2020a) noted concerning the facet approach to principal job satisfaction, this approach might overlook critical facets and the differential importance assigned to the various factors. Hence, in the present investigation, we infer principal self-efficacy as a generalized, multi-item measure of self-efficacy (based on the widely used Schwarz and Hallum, 2008, generalized measure of self-efficacy) rather than a higher-order factor based on multiple facets. (Also see Supplemental Materials, ***SM1***: Bandura's conceptualization of self-efficacy and use of the term in applied and educational research for further discussion).

Relations Between Work Self-Efficacy and Job Satisfaction

A wealth of research shows that job satisfaction and self-efficacy are positively correlated in educational settings (e.g., Skaalvik, 2020b) and organizational settings more generally. (e.g., Judge et al., 1997; 2001). In some studies, the relation was evaluated with regression equations in which self-efficacy (and, sometimes, other predictor variables) was used to predict job satisfaction as an outcome variable.

Results of the regression analysis are sometimes interpreted as if they support a unidirectional model of the directional ordering of self-efficacy →job satisfaction. However, the correlational (cross-sectional) nature of these studies – as often acknowledged by the authors – is unable to differentiate between any of the possible unidirectional models (SE→JS or JS→SE) from each other or from a bidirectional model (SE→JS AND JS→SE), or even a model with not directional ordering (i.e., even though job satisfaction and self-efficacy are substantially correlated at each wave, self-efficacy at time 1 does not predict job satisfaction at time 2 after controlling for job satisfaction at T1, and job satisfaction time 1 does not predict self-efficacy at time 2 after controlling for self-efficacy at T1). Hence, our research fills this gap in the research literature on the nature of the relation between job satisfaction and self-efficacy

The theoretical basis for reciprocal effects between self-efficacy and outcomes stems from Bandura's (1977) theoretical model. However, several alternative theoretical models in applied psychology also posit reciprocal relations consistent with Bandura's model (e.g., Conservation of Resources Theory; Hobfoll, 1989; broaden-and-build theory, Fredrickson 2001, 2008). They posit that resources like self-efficacy and outcomes like job satisfaction are reciprocally related, resulting in positive gain or negative loss spirals (also see (Salanova et al., 2006; Schutte, 2013).

The job-demand resources (JD-R) model (Bakker et al., 2014, 2017) posits self-efficacy as a resource that leads to positive outcomes such as job satisfaction. Thus, job resources are functional aspects of the work setting in achieving work goals, motivating growth, and job satisfaction (the motivation process). In their review of existing research and directions for further research, Bakker and Demerouti (2014) highlighted several longitudinal studies demonstrating what they referred to as reversed causal effects; prior outcomes affect changes in subsequent resources. These reverse causality effects can result in feedback loops between resources, demands, and outcomes that might lead to positive or negative spirals (e.g., Dicke, Stebner, et al., 2018). Dicke et al.'s (2020; 2021) review of this research noted the need for more appropriate longitudinal databases and more robust statistical models of longitudinal cross-lag panel designs. The reverse causality feedback loops noted in JD-R research are simply an alternative way of saying that resources and outcomes are reciprocally related.

In perhaps the only true longitudinal study to study reciprocal relations between job satisfaction and self-efficacy, Wu and Griffin (2012) evaluated reciprocal effects between job satisfaction and a six-item composite measure designed to represent Judge et al.'s core self-evaluations (e.g., self-efficacy, self-esteem, locus-of-control, and neuroticism). In an organizational (non-school) setting, job satisfaction was measured over 10 years, whereas the self-evaluation traits were measured in years 5 and 10. Wu and Griffin used a manifest measure of job satisfaction rather than latent variables because of the model's size. In terms of testing the reciprocal effects, core self-evaluations (Year 5) had a substantial impact on job satisfaction (Year 6, controlling for job satisfaction in Year 5) and subsequent change in job satisfaction (over the Years 6-9). However, job satisfaction (Year 6) and growth in job satisfaction (over the Years 6-9) also had statistically significant effects on core self-evaluations in Year 10. Despite complications in the statistical modeling and interpretational issues associated with the core self-evaluation construct, this is one of the few studies to test reciprocal relations between job satisfaction and a dispositional construct that includes self-efficacy. In contrast to their interpretation of Judge et al.'s (1997; 2001) dispositional perspective, Wu and Griffin argued that it is best to select employees with high core self-evaluations; their results support the potential to improve core self-evaluations by enhancing job performance satisfaction. However, we know of no research testing these implications in an educational setting or even leading in organizational settings.

In summary, a wealth of data shows that job satisfaction and self-efficacy are correlated, and several theoretical models posit that their directional ordering is reciprocal, both for leaders and employees more generally and school principals more specifically. However, appropriate longitudinal research on the reciprocal links between job satisfaction and job self-efficacy is largely lacking. This is particularly so within school leadership research, where we are unaware of any studies examining reciprocal links between these constructs among leaders. We also note serious methodological limitations in the design and analysis of existing research and a groundswell of new methodological approaches to address these issues that have not been considered in studies of the relations between job satisfaction and self-efficacy. Thus, we now move to a focus on methodological issues underpinning our research.

Methodological Emphasis: Reciprocal Effects & Models of Cross-lagged Panel Models

Cross-Lagged Panel data

In cross-lagged panel models (CLPMs), two or more variables (job satisfaction and self-efficacy) are measured repeatedly over time. CLPMs test reciprocal effects between the variables over time. The cross-lag aspect refers to the focus on effects leading from one variable to another over time (i.e., effects of prior job satisfaction and self-efficacy on future job satisfaction and self-efficacy). The critical concern is the directionality of effects and directional predominance (i.e., whether the directional effects are stronger for one construct than the other). Although it is possible to test some models with only two waves, more waves are desirable. Particularly when there are at least three waves, it is possible to test the consistency of effects over time for the different variables.

Typically, cross-lagged panel studies consider only lag-1 effects (i.e., paths relating variables in adjacent waves). Indeed, as many cross-lagged panel studies consider only two waves of data so that only lag-1 effects can be estimated. However, when there are more than three waves, it is possible to consider paths between non-adjacent paths (i.e., Lag 2 effects). Indeed, Marsh, Pekrun, et al. (2018) argued that lag-2 effects should be considered. They suggested the critical issue was whether or not critical interpretations concerning research questions varied due to the inclusion of lag-2 paths. They found that lag-2 cross-lagged paths were mostly small and their inclusion had little effect on substantive interpretations. Lag-2 paths might have an a priori substantive interpretation, but also provide a stronger control for pre-existing differences. Hamaker and colleagues (e.g., Hamaker et al., 2015; Mulder & Hamaker, 2021) further noted that CLPMs often have to include lag-2 effects to achieve an acceptable fit comparable to RI-CLPMs' fit. Because RI-CLPMs include a global (between-person) trait factor based on measures across all the waves, it is unlikely that lag-2 effects for within-person stability effects will substantially improve fit associated with pre-existing differences. Based partly on Marsh, Pekrun et al., Lüdtke and Robitzsch (2021) noted the importance of including lag-2 effects in CLPMs to control potentially confounding covariates.

Distinguishing Between RI-CLPM (Within-Person) and CLPM (Between-Person) Perspectives

Historically, CLPMs are the most widely used model in psychological research to test directional ordering based on panel data. However, following Hamaker and colleagues (Hamaker, 2018; Hamaker et al., 2015; Hamaker & Muthén, 2020; Mulder & Hamaker, 2021), the popularity of RI-CLPMs has surged. However, Orth et al. (2021) stressed that these two models address different questions, result in different interpretations, and make different assumptions. Here we argue that both models separately and their juxtaposition are relevant to different theoretical questions. Hence understanding these differences is critical (See ***SM2*** for further discussion),

For both CLPMs and RI-CLPMs, the critical parameters are stability paths (Bxx and Byy in Figure 1), and particularly the cross-lagged paths (Bxy and Byx in Figure 1). If both Bxy and Byx are statistically significant, X and Y are said to be reciprocally related. If both these cross-lagged paths are statistically significant but differ significantly in size, one of the constructs is directionally predominant. For example, if paths from prior self-efficacy to subsequent job satisfaction were greater than paths from prior job satisfaction to self-efficacy, self-efficacy would be said to be directionally predominant in job satisfaction.

The RI-CLPM includes a stable trait factor (Tx and Ty in Figure 1), but the CLPM does not. Hence, CLPMs are nested under the RI-CLPM. In this sense, RI-CLPMs model how scores obtained at each wave differs from a person's stable trait (a within-person perspective disaggregated from the between-person perspective reflected in the stable trait). Furthermore, these time-specific within-person variations observed at one point in time influence within-person-variations observed at later time points. In contrast, CLPMs model how inter-individual differences observed at one point in time are related to inter-individual differences observed at later time points (a between-person perspective).

The critical difference in distinguishing between these models is in how the term between-person is used. In the CLPM, the (undecomposed) between-person effects reflect a combination of deviations from a global trait (within-person effects) and a stable trait (decomposed between-person effects). This use of the term between-person effects is consistent with its use in most studies of individual differences in relations among variables and in most cross-sectional studies. In contrast, the RI-CLPM decomposes these effects into separate within- and between-person components. The use of the generic term between-person is appropriate within the context of each of these two models. However, to highlight this distinction and avoid confusion, we use the terms “decomposed between-person effects” (RI-CLPM) and “undecomposed between-person effects” (CLPM).

The potential confounding of effects with unmeasured covariates threatens the interpretation of both CLPMs and RI-CLPMs. The RI-CLPM’s key advantage is providing increased control for time-invariant (between-person) covariates not included (e.g., Hamaker et al., 2015; Mulder & Hamaker, 2021; Marsh, Pekrun et al., 2022; Pekrun et al., 2022). The rationale for this claim is the global trait factor is intended to absorb the time-invariant confounders so that the trait factor represents these confounders. These global train effects are statistically independent of the within-person autoregressive factors. Thus, unmeasured covariates that are truly time-invariant are most likely to influence the sizes of the (decomposed between-person) global trait factors, but are unlikely to affect the within-person autoregressive factors used to evaluate directional ordering. This distinction is important in interpreting the results, particularly tests of directional ordering and directional predominance. Of course, it is possible to include measured covariates into RI-CLPMs; Mulder and Hamaker (2021) described how to do this with manifest RI-CLPMs. Although CLPMs can also control measured covariates, they provide less control for the confounding effects of unmeasured covariates.

Neither RI-CLPMs nor CLPMs provide strong control for unmeasured covariates that vary over time (or time-invariant covariates whose effects vary from wave to wave, possibly reflecting an unmeasured process). The best way to control the effects of unmeasured (time-varying and time-invariant) covariates is to include them in the study's design and incorporate them into the statistical models. Hence the selection of covariates is crucial for RI-CLPMs and particular CLPMs (also see VanderWeele, 2019). However, following Marsh, Pekrun et al. (2018; also see Marsh, Pekrun et al., 2022; Marsh, Dicke, et al., 2002), Lüdtke and Robitzsch (2022) emphasized the inclusion lag-2 effects to provide more robust controls for confounding (VanderWeele et al., 2020). Furthermore, the goodness-of-fit of CLPMs with lag-2 effects variables was similar to the fit of RI-CLPMs. Thus, goodness-of-fit is no longer a critical issue in the comparison of RI-CLPMs and CLPMs with lag-2 effects. Following Lüdtke and Robitzsch's recommendations, we extended traditional CLPMs based on lag-1 estimates to include lag-2 estimates.

In summary, RI- CLPMs and CLPMs address distinct questions and typically have different or even contradictory interpretations. Their comparative strengths, weaknesses, and appropriate interpretations are the basis of ongoing controversy and debate. Hence, educational and psychological researchers need to understand these differences in CLPMs and RI-CLPMs. Our study is the first to juxtapose their theoretical rationale and results in tests of the directional ordering of job satisfaction and self-efficacy. We aim this presentation to educational psychology researchers based on a classic issue that is surprisingly understudied. Further, we demonstrate extensions of RI-CLPMs, establish the benefits of lag-2 effects, and show their usefulness in educational and psychological research.

The Present Investigation: A Priori Predictions and Research Questions

Based on our empirical and theoretical literature review, we posit research hypotheses (where there is a clear a priori basis for directional predictions). We also propose four research questions and relevant discussion for critical issues where there is no basis for a priori predictions.

Research Hypothesis 1: Directional-Ordering - Model of Reciprocal Effects

We predict a priori that leaders’ job satisfaction and self-efficacy will be reciprocally related; both constructs in one wave will significantly predict both constructs in the subsequent wave (Figure 1). Unfortunately, there is surprisingly little empirical research relevant to this issue based on appropriate designs and analyses. However, our predictions are consistent with Bandura's reciprocal principle for self-efficacy, theoretical extensions of the JDR and COR models (see Salanova et al., 2010), broaden-and-build theory (Fredrickson, 2001; 2008), Wu and Griffin's (2012) empirical results, and extensive research showing that self-beliefs and outcomes are reciprocally related to other outcomes (e.g., Huang, 2011; Marsh, Pekrun et al., 2022; Marsh & Craven, 2006; Valentine, et al., 2004).

Four Research Questions: Extensions of RI-CLPMs and Juxtaposition with CLPMs

Mulder and Hamaker (2021; Hamaker et al., 2015) noted, as have many others, that there was no a priori basis for predicting how effects based on CLPMs and RI-CLPMs should differ. However, based on simulated data, they offered valuable extensions to the traditional RI-CLPM. These extensions included the specification of a fully latent model with multiple indicators of each trait and the addition of person-level covariates based upon fully manifest models. In juxtaposing RI-CLPMs and CLPMs, Lüdtke and Robitzsch (2021) argued the importance of lag-2 paths based on simulated data with manifest models. However, neither the addition of covariates by Mulder and Hamaker nor the addition of lag-2 effects by Lüdtke and Robitzsch (2021) were based on fully latent models. Mulder and Hamaker presented tests of the invariance and discussed the invariance of the measurement model for their fully latent model based on simulated data. However, they did not extend this to the invariance of the structural aspect of the model (e.g., stability and cross-lag paths) and tests of stationarity. Extending these methodological studies, we apply and further develop these additional features in relation to our research hypothesis that job satisfaction and self-efficacy are reciprocally related. In our discussion of these extensions, we posit five research questions

Research Question 1: Juxtaposing RI-CLPM and CLPMs Interpretations. Our key research question is whether support for the predicted reciprocal effects (Research Hypothesis 1) differs for RI-CLPMs and CLPMs. Nevertheless, Hamaker et al. (2015; Mulder & Hamaker, 2021; also see earlier discussion) note that there is no a priori basis for predicting how RI-CLPM and CLPM estimates differ in size or even direction. We know that the two models are equivalent if between-person global traits in RI-CLPMs (i.e., the Tx and Ty in Figure 1) have zero variance (e.g., Hamaker et al., 2015). Hence the underlying rationale for applying RI-CLPMs is the presence of a substantial between-person global trait component (i.e., the Tx and Ty in Figure 1). Based on the substantial test-retest stability of the constructs (e.g., Dicke, Marsh, et al., 2018; also see subsequent discussion of Table 1), the global trait factors in the RI-CLPM should be substantial. Nevertheless, we begin by evaluating the size of global trait factors as a preliminary basis for applying RI-CLPMs. Because global trait factors capture the long-term stability in RI-CLPMs, the stability paths are expected to be substantially smaller for RI-CLPMs than CLPMs. However, there is no a priori basis for predicting differences in the critical cross-paths in the two models.

Research Question 2: Tests of Causal Predominance. Historically, there was a focus on directional-ordering predominance (e.g., Calsyn & Kenny, 1977; also referred to as causal predominance). However, this focus has been largely dropped in applying CLPMs and RI-CLPMs. Thus, even when there is support for reciprocal effects, one set of cross-paths (e.g., Bxy in Table 1) might be significantly greater than the other (e.g., Byx). Also, tests of statistical significance might support a unidirectional model (e.g., Bxy significant, but not Byx). However, if the difference between Bxy and Byx is not statistically significant (i.e., there is no directional-ordering predominance), then claimed support for a unidirectional model would be problematic. In studies of job satisfaction and self-efficacy, Wu and Griffin (2012) characterized much previous research as based implicitly on a unidirectional model (self-efficacy →job satisfaction) rather than the bidirectional model posited by Wu and Griffin, and in the present investigation. However, a compromise between the two might be a bidirectional model with directional-ordering predominance of the self-efficacy →job satisfaction path over the job satisfaction→self-efficacy path. Because formal tests of directional ordering predominance are relevant to the present investigation and CLPMs and RI-CLPMs more generally, we demonstrate their application and implications for the interpretations of our results.

Research Question 3: Extended Models: lag-2 effects. Following Lüdtke and Robitzsch (2021) and Marsh, Pekrun et al. (2018; 2022), we extended CLPMs and RI-CLPMs to include lag-2 effects. To the extent that lag-2 effects are consistent over waves, these effects are likely to be absorbed into the global trait factors in RI-CLPMs (i.e., the Tx and Ty in Figure 1). If this is the case, the inclusion of lag-2 paths should substantially improve CLPM's fit and reduce the size of CLPM's lag-1 stability paths, but have less effect on these values for RI-CLPMs. However, the critical question is how the inclusion of lag-2 paths will influence the cross-paths and the hypothesized support for the reciprocal effects (Research Hypothesis 1).

Research Question 4: Extended models: covariates and invariance constraints. For both RI-CLPMs and CLPMs, we evaluated extended models that included controls for covariates: gender, age, position (school principal or other school leaders), and sector (public or private). Relations between these covariates and leaders' job satisfaction and self-efficacy are substantively interesting. However, our main concern is how their inclusion affects the cross-lagged paths. In this respect, we view the inclusion of covariates as a sensitivity analysis concerning interpretations of our results and assumptions of the models.

Because we test fully latent models, we can explore invariance issues not previously considered with CLPMs and RI-CLMPs. In pursuit of this extension, we discuss the invariance of the measurement model that has been widely considered in CLPMs and for simulated data in the Mulder and Hamaker (2021) latent RI-CLPMs. We extend this discussion to include longitudinal invariance, including formal tests of stationarity that apparently have not been applied to either CLPMs or RI-CLPMs. In each step of this process, the overarching question is to what extent there is support for the generalizability of interpretations (particularly Research Hypothesis 1) concerning the invariance constraints and the implications of failure to support invariance. We discuss these issues in detail (see extended discussion in SI:3) because these issues are not well understood and are largely unconsidered in studies of the relations between job satisfaction and self-efficacy.

Method

Sample

Participants were a large (*N* = 5663) national representative sample of Australian school principals (74%), deputy principals and other school leaders (26%) who were surveyed between 2011 and 2019. The school leaders worked in primary schools (64%), secondary schools (22%), and "other" settings (14%; e.g., Kindergarten -Year 12 schools, special education schools). The average age was 57.6 (SD 7.3) and 56% of the participants were women. On average, participants had been in their current leadership position for 5.2 years (SD = 4.3) and in leadership roles generally for 12.5 years (SD = 7.3).

Participants responded to invitations sent by national and state-based school principal organizations inviting them to complete our annual survey each year, 2011 and 2019. Nearly half the school principals in Australia are included in this sample. The schools and their principals are broadly representatives of the population of Australian school principals as a whole (see https://www.XXX). As a consequence of this sampling design, there are overlapping sets of school leaders who completed the survey each year, with some previous participants dropping out and new participants joining the survey. Also, school leaders who missed one year occasionally completed surveys in subsequent years. Thus, our study's total number of leaders was 5,663, but the number who completed surveys each year varied (2019 = 1894; 2018 = 2289; 2017 = 2549; 2016 = 2841; 2015 = 2641; 2014 = 2467; 2013 = 2010; 2012 = 2084; 2011 = 2049).

Our study is part of on an overarching research program entitled " School Principals Diminishing Wellbeing: What Makes A Positive Difference? " funded by XXX and led by lead principal investigator XXX. Our University Ethics Review Committee judged this research program (including this study) to be low-risk, approved its undertaking (Ethics report number xxx), and subsequently extended approval through 2025.

Measures

Job satisfaction and self-efficacy measures were from the Copenhagen Psychosocial Questionnaire. This survey has been used in thousands of enterprise-based assessments of psychosocial risk in the workplace (Nübling et al., 2014; Pejtersen et al. (2010). Based on a total of 34 factors, the Copenhagen Psychosocial Questionnaire covers a broad array of key workplace factors based on the leading measure and theories of occupational health and wellbeing. Furthermore, this instrument and its factors have been validated for Australian school leaders (see Dicke, Marsh, et al., 2018).

Job satisfaction was based on responses to four items (How pleased are you with the following: your work prospects; the physical working conditions; the way your abilities are used; and your job as a whole, everything taken into consideration?) on a four-point response scale (Very satisfied, Satisfied, Unsatisfied, Very unsatisfied).

The self-efficacy scale in the Copenhagen Psychosocial Questionnaire is classified as a personality variable reflecting a generalized trait. It contains six items selected from Schwarzer’s (Schwarzer & Jerusalem, 1995) General Self-Efficacy scale, using the same item-wording and Likert-type response scale as the Schwarzer instrument. The Schwarzer (Schwarzer & Jerusalem, 1995) General Self-Efficacy scale is one of the most widely used measures of generalized self-efficacy, having been translated into more than 32 languages with good psychometric properties (unidimensionality and reliability) across countries (e.g., Luszczynska et al., 2005; Sherer et al., 1982; Sniehotta et al., 2005). Respondents to the measure of self-efficacy on the Copenhagen Psychosocial Questionnaire are asked: How well do these descriptions fit on you as a person? (I am always able to solve difficult problems, if I try hard enough; If people work against me, I find a way of achieving what I want; It is easy for me to stick to my plans and reach my objectives; I feel confident that I can handle unexpected events; When I have a problem, I can usually find several ways of solving it; Regardless of what happens, I usually manage) on a four-point response scale (Fits perfectly; Fits quite well; Fits a little bit; Does not fit).

We note that the wording of the items reflects a general measure of self-efficacy like that typically assessed in organizational psychology studies (e.g., Chen et al., 2001; Judge et al., 1997; Locke et al., 1996; Schwarzer & Jerusalem, 1995). However, because the overarching survey and the items from the Copenhagen Psychosocial Questionnaire focus specifically on the work context in a school setting, respondents responded to these items concerning their work context. Thus, for example, before responding to any items, respondents were told: *The survey is being conducted in response to concerns that the increasing complexity and workload demands of school leadership roles are impacting on the health and wellbeing of Australian school leaders.*

In factor analyses of all 34 factors (Dicke, Marsh, et al., 2018) measured by the Copenhagen Psychosocial Questionnaire, both the job satisfaction and self-efficacy constructs were well-defined, reliable, and unidimensional: self-efficacy (CFA factor loadings of .453 - .742) and job satisfaction (CFA factor loadings of .516 - .850). Dicke, Marsh, et al. also reported support for the convergent and discriminant validity of job satisfaction and self-efficacy in relation to time and to the other 32 work-related constructs included in the COPSOC (e.g., health, burnout, stress, job security, role conflict, social support, meaning, trust) and related correlates (e.g., depression, autonomy, confidence, sources of stress). Consistently with a priori predictions and supporting convergent validity, self-efficacy was most highly correlated with confidence (r = .449); in support of discriminant validity, the stability of self-efficacy over one year was substantially higher (r = .635) than the correlations between self-efficacy and confidence within each wave.

Covariates

We also included several individual-difference covariates (also see discussion of Research Question 4): gender, age (birth year—linear and quadratic components), leadership role (school principal or other—typically deputy principal or head teacher), school sector (public vs. catholic and independent), and school level (primary vs. secondary). Extending the Murder and Hamaker (2021) specifications for manifest RI-CLPMs, there are several different alternatives such that the covariates can directly affect the: (1) observed variables (the unlabelled boxes in Figure 1); (2) latent traits representing the time-specific latent factors at each wave (the Xs and Ys in Figure 1), or (3) global latent trait factors (Tx and Ty in Figure 1). Here we focus on the second two of these alternatives (that are most similar to the corresponding CLPMs).

Statistical Analyses

We used Mplus 8 (Muthén & Muthén, 2008-19) and the robust maximum likelihood estimator for all analyses. We included all participants who responded to at least one wave. There are critical limitations in the use of traditional approaches to missing data, particularly in longitudinal studies (Graham, 2009). Therefore, we used the full information maximum likelihood method (FIML; Enders, 2010) to make maximum use of cases with missing data (***SM6*** for further discussion). FIML provides trustworthy, unbiased estimates even when there is substantial missing data (Enders, 2010), particularly in large longitudinal studies (Jelicić et al., 2009).

For all models, we evaluated goodness-of-fit with traditional indices that are reasonably independent of sample size (CFI = confirmatory fit index, TLI = Tucker-Lewis index, and RMSEA = root mean square error of approximation) as well as the chi-square. TLI and CFI vary along a 0-to-1 continuum in the population; values greater than .90 and .95, respectively, reflect acceptable and excellent fits to the data, respectively. Values of RMSEA smaller than .08 and .06 support acceptable and good model fits. For the differences between two nested models, if the decrease in fit for the more parsimonious model is less than .01 for CFI and TLI, then the more parsimonious model is supported. RMSEA and the TLI also incorporate a penalty for lack of parsimony so that it is possible for more restrictive models to better fit less restrictive models. However, it is critical to emphasize that these recommended cut-off values are only rough guidelines and do not constitute "golden rules" (Marsh et al., 2004).

Preliminary Analyses

We tested a series of CFA measurement models based on responses to 90 items (4 job satisfaction and 6 self-efficacy items in each of 9 waves). For these models, we merely specified two latent variables (job satisfaction and self-efficacy) for each wave. We allowed all latent correlations (within-wave and over time) to be freely estimated. In order to facilitate interpretations, we standardized all items for job satisfaction and self-efficacy to a common metric based on Wave-1 responses (i.e., Mn = 0, SD = 1 for 2011, wave 1 of the study). We evaluated goodness-of-fit with indices (CFI, TLI, and RMSEA) that are relatively independent of sample size.

When the multiple indicators of each construct are parallel over the different time waves, testing the invariance of the measurement structure over time is critical. In fully-latent models of longitudinal data, it is typical to evaluate a set of models with different invariance constraints of parameters over time (e.g., Marsh, Morin et al., 2014; Meredith, 1993; Millsap, 2011; Mulder & Hamaker, 2021): configural, with no invariance constraints; metric, with factor loading invariance; and scalar, with intercept invariance. Following Marsh et al. (2013; for further discussion, see Marsh et al., 1996; Joreskog, 1979), we also tested correlated uniquenesses relating residual variance terms for the same item in different waves. Failure to include them typically results in biased parameter estimates and a poorly fitting model.

In preliminary analyses, we tested whether the measurement model is well-defined. These tests are not based on any particular model (e.g., CLPM or RI-CLPM). However, applying either CLPMs or RI-CLPMs models is dubious unless there is reasonable support for at least configural invariance. Furthermore, if support for the invariance of factor loadings is weak, then invariance tests for other associated parameters (e.g., stability and cross-lagged paths in CLPMs and RI-CLPMs) are questionable. This is a critical issue in that many cross-lagged panel studies are based on manifest models that preclude tests of the measurement model (see extended discussion in Supplemental Materials, ***SM2, SM3, SM4***).

As expected, measurement model MM0 (Table 1) with no correlated uniquenesses provided a poor fit to the data (see earlier discussion). Measurement model MM1 with configural invariance included correlated uniquenesses but imposed no invariance constraints; it had a very good fit to the data (RMSEA = .011, CFI = .980, TLI = .975; Table 1). Model MM2 (metric, with factor loading invariance) also had a good fit to the data (RMSEA = .011, CFI = .979, TLI = .975). MM3 (scalar, with intercept invariance) resulted in a slightly poorer fit (RMSEA = .012, CFI = .974, TLI = .970), but one that was still very good in relation to traditional guidelines (e.g., Marsh et al., 2004). These preliminary analyses support the need to include correlated uniquenesses (MM1 vs. MM0), and the invariance of factor loadings (MM 2 vs. MM1). There is reasonable support for intercepts (MM3 vs. MM2). We based our primary analyses on Model MM2 (invariance of factor loadings--metric invariance) as the invariance of intercepts is not relevant for covariance structures models considered here.

All solutions were identified by fixing the factor loading of each construct's first indicator to a constant value. However, to facilitate interpretations, instead of fixing the value to 1.0, we fixed it to the standardized factor loading in the metric invariance solution. The resulting solution then has a factor variance of 1.0 in wave 1. However, the factor variances in subsequent waves can vary relative to the factor variance in wave 1. Thus, parameter estimates are standardized in relation to a common metric (that facilitates the comparison of parameter estimates in different waves), resulting in an unstandardized solution that approximates the completely standardized solution. This parameterization is also useful in subsequent analyses where we test differences in the cross-lagged paths (Bxy and Byx in figure 1).

Importantly, all CLPMs and RI-CLPMs are nested under our measurement model MM2 (see End Note 1). Hence, MM2 is a critical basis of comparison for CLPMs and RI-CLPMs, including structural invariance constraints and extensions to include additional lagged parameters. In particular, relations among all 18 factors (job satisfaction and self-efficacy over the 9 waves) are freely estimated. Thus, MM2 is completely saturated in terms of relations among the 18 factors. In contrast, CLPMs and RI-CLPMs and their extensions all place constraints upon these relations. Thus, the fit of the constrained CLPMs and RI-CLPMs in relation to the measurement model MM2 provides a global test of these constraints imposed by the CLPMs and RI-CLPMs. This is a critical contribution, because the fit of the CLPMs and RI-CLPMs is rarely interrogated in relation to the measurement model—even for fully-latent CLPMs and RI-CLPMs.

Results

Correlations Among the Variables

We begin by evaluating the latent correlation matrix (Table 2) of relations between the 18 latent factors (job satisfaction and self-efficacy measured the 9 waves). This represents a fully latent multitrait-multimethod (MTMM) correlation matrix with “time” as the method factor (see Marsh, Huppert, et al., 2020; Marsh, Lüdtke, et al., 2010). Thus, the results indicate that both constructs are highly stable and consistent over the nine waves—consistent with the logic of the RI-CLPM. The average of lag-1 correlations (i.e., test-retest correlations for constructs measured in adjacent annual waves) for matching traits is .68 (.67-.72) for self-efficacy and .67 (.64-.70) for job satisfaction. Consistently with the characteristic simplex pattern, lag-2 test-retest correlations are somewhat smaller (M r = .60 for self-efficacy and .56 for job satisfaction). However, even the estimated correlations for measures in 2011 and 2019 are substantial (.52 for self-efficacy and .55 for job satisfaction). These results demonstrate substantial stability over time for both self-efficacy and job satisfaction.

Correlations between job satisfaction and self-efficacy within each of the nine data waves vary from .34 to .46 (mean r = .40). Furthermore, correlations remain substantial for job satisfaction in one wave with self-efficacy in the next wave (mean r = .32 for lag-1 correlations) and self-efficacy in one wave with job satisfaction in the next wave (mean r = .33). Although the two constructs are moderately correlated, there is good evidence that they are well-differentiated. Thus, correlations among self-efficacy factors and among job satisfaction factors are substantially higher than any correlations between self-efficacy and job satisfaction. These results demonstrate that job satisfaction and self-efficacy are substantially correlated within each wave and from one wave to the next.

Also of interest are the relations between covariates and the self-efficacy and job satisfaction factors (Table 2). Even though many are statistically significant due to the large sample size, these relations tend to be small. Furthermore, the small correlations are reasonably consistent over the nine data waves. The largest correlations with the covariates are for job satisfaction, which are higher for school leaders in the private sector. Both job satisfaction and self-efficacy are marginally higher for school leaders who are school principals (rather than other school leaders -- deputy principals and heads of departments), male school leaders, and older school leaders (non-linear effects of age are non-significant).

CLPM: Directional Ordering

We evaluated support for sets of invariance constraints for the basic CLPMs (Tables 3 and 4; MB1, longitudinal equilibrium; MB2, partial-stationarity). We also tested models constraining the cross-lagged paths to be equal (Bxy = Byx) as a test of directional predominance for each model (see End Note 2). In addition, we also tested a model with no invariance constraints (MB1 in Table 3) to provide a basis of comparison for the fit of the other six models. Across all seven models, goodness-of-fit is remarkably consistent (e.g., TLIs are .958 for all seven models, Table 3). Particularly relevant is the finding that the model with no invariance over the nine waves for any autoregressive stability (Byy, Bxx) and cross-lagged (Bxy, Byx) paths had a similar fit to the modelsthat constrained all these coefficients to be equal over the nine waves (MB models with constraints in Table 3). These results support full-trend stationarity of parameter estimates over the nine waves.

The critical parameter estimates for the CLPMs (see Figure 1) are the stability paths (Bxx and Byy) and particularly the cross-lagged paths (Bxy and Byx). We constrained each of these four paths to be invariant (equal) across the nine data waves for all models but model MB1. The stability paths are substantial for both job satisfaction and self-efficacy (varying from .678 to .713, Table 4). The cross-lagged paths are significantly positive but much smaller in size (varying from .055 to .065). There is no support for the directional predominance of either self-efficacy or job satisfaction. Indeed, the difference between the two cross-lagged paths was not statistically significant (see endnote 1) for models MB1 and MB2 (Table 2).

The substantive interpretation of results for the CLPM are clear-cut. There are small, but highly consistent reciprocal effects between job satisfaction and self-efficacy. Furthermore, there is no evidence that either of these constructs is directionally predominant over the other. Methodologically, the results are highly consistent over models positing alternative sets of invariance constraints (including tests of full-trend stationarity that have not previously been investigated with CLPMs—see **SM3**).

RI-CLPM: Directional Ordering

For the RI-CLPMs (Table 3), we evaluated models of structural invariance and directional predominance models that were parallel to the CLPMs. Again, goodness-of-fit is remarkably consistent (e.g., TLIs varied from .974 to .975, Table 3). Similarly, in tests of directional predominance, none of the differences between the two cross-lagged paths was statistically significant.

For the RI-CLPMs, we evaluated stability and cross-lagged paths for the same set of models considered for CLPMs (Table 4 and Figure 1). The major difference, of course, is that the RI-CLPMs have global trait factors (Tx and Ty in Figure 1) for both self-efficacy and job satisfaction. The variance components (Table 4) for the global trait factors are consistently substantial for both self-efficacy (.531-.536, TrtVarSE inTable 4) and job satisfaction (.589-.661, TrtVarJS). Thus, the stable (decomposed between-person) trait factors explained slightly more than half of the variance in both constructs. The correlation between the two global trait factors is also consistently substantial (TrtCov SE-JS in Table 4; .477-.481). We note that the global (decomposed between-person) trait correlation estimates are consistently higher than any latent correlations between traits within each of the nine waves in the basic latent measurement model (.34 to .46; mean r = .379 in Table 2).

The primary focus of the RI-CLPM, like the CLPM, is the two stability coefficients (Bxx and Byy) and particularly the two cross-lagged paths (Bxy and Byx). The stability paths are significantly positive for both self-efficacy (.270-.280) and job satisfaction (.198-.227), as are the correlations between these two factors (.231-.286; Table 4). All the cross-lagged paths are positive but much smaller (varying from .035 to .066). All cross-lagged paths are statistically significant for RI-CLPMs with longitudinal equilibrium and partial invariance constraints. Furthermore, when the two cross-lagged paths were constrained to be equal, the standard errors were substantially smaller. These models demonstrate that both cross-lagged paths are significant and equivalent under conditions of longitudinal equilibrium and partial stationarity

Across the RI-CLPMs with the cross-lagged paths freely estimated, the sizes of cross-lagged paths from self-efficacy to job satisfaction appear to be marginally higher than those from job satisfaction to self-efficacy (see Research Question 2). However, the difference is not statistically significant for any of the three models with invariance constraints (all ps > .10). Hence there again is no support for the directional predominance of either construct over the other.

The substantive interpretation of the RI-CLPM results is also straightforward. Both job satisfaction and self-efficacy have a substantial stable trait component. There are small, but highly significant reciprocal effects between the within-person factors for job satisfaction and self-efficacy, but no evidence that either construct is directionally predominant over the other. Methodologically, the results and goodness-of-fit indices are highly consistent over different invariance constraints. More broadly, the methodological implications are potentially important. This is one of a relatively few studies to test fully-latent RI-CLPMs and the first to test full-trend stationarity.

Extended Models for CLPMs and RI-CLPMs

Starting with model MB2 (Table 3), we extend the basic CLPMs and RI-CLPMs in several ways. First (Models ML1-ML3 in Tables 3 and 5), we added lagged paths linking non-adjacent waves (lag-2 paths; e.g., paths leading from factors in wave 1 to factors in wave 3), particularly for the autoregressive stability paths (see discussion of Research Question 3). Second (Models MC1-MC5 in Tables 3 and 5), we included covariates in the models (see Research Question 4). Finally, we evaluated models with both additional lagged paths and covariates (Models MLC1-MLC5 in Tables 3 and 5).

CLPM: Additional Lags

Because the corresponding basic model with no covariates (MB2 in Table 3) is nested under these models, it is appropriate to compare fit indices for models containing additional lags (ML1-Ml3 in Table 3) with the fit of the basic model MB2. Consistently with expectations, the additional lags improved the fit of the CLPMs (Table 3). Compared to the fit of the lag-1 model (TLI = .958, MB2 in Table 3), the CLPM with lag-2 paths was better (TLI = .971, MML1 in Table 3). This improved fit was largely due to the stability paths, as elimination of the cross-lagged paths had little effect on fit (TLI = .971, ML2 in Table 3). Finally, the addition of Lag-3 stability paths led to some further improvements in fit (TLI = .974, ML3 in Table 3). Indeed, the fit of ML3 (Table 3) is only marginally lower than the fit of the corresponding measurement model (TLI= .975, MM2 in Table 1) that is fully saturated in terms of structural model constraints used to represent relations among factors in CLPMs. This comparison provides additional support for including lag-2 and even lag-3 paths in CLPMs.

How does the inclusion of additional lags affect the key parameter estimates? Compared to the corresponding lag-1 model (MB2 in Table 3), autoregressive stability and cross-lagged path coefficients for models with additional lags (ML1-ML3, Table 5) were smaller. As expected, these differences were most evident in the stability coefficients (i.e., lag-1 effects were much smaller). Thus, for example, the lag-1 stability paths in ML3 that included lag-2 and lag-3 paths (.404 and .383 for job satisfaction and self-efficacy, respectively) were substantially smaller than the corresponding values in MB2 that had only lag-1 paths (.679 and .713). Likewise, whereas there were also small differences in the cross-lagged paths (.041 and .048 vs. .059 and .062, respectively), these differences were much smaller. Importantly, however, the cross-lagged paths remained all statistically significant, and tests of directional predominance (i.e., differences between the cross-lagged effects for each construct) remained all non-significant. Hence, in this respect, the interpretation of these cross-lagged paths did not change even though the values were slightly smaller.

CLPM: Covariates

Models that include covariates are not nested under the models not including covariates that we have considered thus far. Hence, the corresponding fit indices are not directly comparable. Thus, we constructed nested models that provide a better basis for comparison. More specifically, we estimated a model in which paths relating the five covariates to the job satisfaction and self-efficacy factors at all nine waves were constrained to be zero (MC4 in Table 3, TLI = .950). When we freely estimated the effects of these covariates, the fit improved – but not a lot (MC1 in Table 3, TLI = .953). Consistently with expectations (Research Question 4), we showed that much--but not all--of the covariates' effects were explained by paths to just the first of the first two waves (MC3 and 4 in Table 3, both TLIs = .952). Thus, the covariates' effects were not substantial, consistent with our earlier discussion of them (Table 2). Consequently, it is unsurprising that the inclusion of covariates had almost no effect on autoregressive stability and cross-lagged path coefficients (those for Models MC1-MC in Table 5 compared to those for MB2 in Table 4). However, the covariates did have small effects beyond the first wave of data, in contrast to typical representations of covariate effects in CLPMs.

CLPM: Additional Lags and Covariates

Because the models with only covariates (the MC models in Table 3) are nested under corresponding models with covariates and additional lags (MLC1-MLC7 in Table), it is reasonable to compare fit indices for the two sets of models. Consistently with earlier discussion, including Lag-2 and Lag-3 stability coefficients improved the fit (e.g., TLI = .966, MLC1 for Lag-1 and TLI = .967, MLC5 for Lag-2 compared to TLI = .952 for MC2). Compared with models where paths from the covariates were constrained to zero, the results showed small effects of the covariates. However, the changes in fit were small (e.g., TLI = .966 for MLC1 vs. TLI = .964 for MLC4, and TLI = .969 for MLC5 vs. TLI = .967 for MLC64). These results are consistent with our expectations that the inclusion of lag-2 and lag-3 stability coefficients would substantially reduce the effects of covariates.

Given the small effects of the covariates, it is not surprising that parameter estimates for models with additional lagged paths and covariates (MLC1-MLC7 in Table 5) are similar to the corresponding models with only additional lagged effects already discussed (ML1-ML3 in Table 5). Likewise, the four key (stability and cross-lagged) paths were very similar for lag-2 and lag-3 models with and without covariates (e.g., ML2 vs. MLC1 and ML3 vs. MLC6). In summary, the covariates had relatively little effect, particularly after including the additional lagged paths.

Extended Models for RI-CLPMs. **W**e also estimated extended models to the basic RI-CLPMs that paralleled those for the CLPMs. However, the interpretations of these extended RI-CLPMs are fundamentally different. By design, the (decomposed) between-person trait factors in RI-CLPMs are meant to represent all the time-invariant (between-person) effects. Based on this rationale, we expected that including covariates and additional lags would have little effect on fit or the (within-person) stability and cross-lagged paths that are of primary interest. Consistently with expectations, the cross-lagged and stability coefficients were little affected by adding covariates and additional lagged effects. These results demonstrate a key advantage of basic RI-CLPMs compared to basic CLPMs in terms of interpretation robustness.

The additional lagged paths for RI-CLPMs only marginally improved the fit (Table 3; TLI = .974 for MB2 with no lagged effects vs. .975, .975, and .976 for ML1-ML3 with lag-2 and lag-3 effects). Likewise, the addition of covariates had little effect on fit. Furthermore, the fit of models with both covariates and additional lag-2 and lag-3 effects (MCL1 – MCL7) had fit indices similar to corresponding models with only lag-2 and lag-3 effects (ML1-ML3 in Table 3). In summary, consistent with expectations, the inclusion of covariates and additional lagged paths had little impact on the fit of the RI-CLPMs. This demonstrates that the random intercepts do well at capturing the trait-like stability of these constructs over time.

Across the RI-CLPMs, parameter estimates in models with additional lagged effects and covariates (Table 5) are similar to those for MB2 (Table 4) with none of these additions. Of particular interest are the (within-person) cross-lagged paths relating prior self-efficacy to subsequent job satisfaction, and prior job satisfaction to subsequent self-efficacy. For example, in model MB2 (Table 4), paths from self-efficacy to job satisfaction (.039) and paths from job satisfaction to self-efficacy (.066) were statistically significant and small, but did not differ significantly from each other (paths = .056 when constrained to be equal). Across the 15 extended models (Table 5), paths from self-efficacy to job satisfaction (.037 - .072) and paths from job satisfaction to self-efficacy (.053 -.072) were consistently significant and small. Importantly, differences between the two cross-lagged paths were not significantly significant (see endnote 2) for any of the models (all ps > .10). In summary, consistent with expectations, including covariates and additional lagged paths had relatively little effect on parameter estimates or substantive interpretations of key parameters.

Discussion

In educational and applied organizational psychology, job satisfaction and self-efficacy are widely studied. Nevertheless, our literature review revealed no fully adequate previous research that evaluated their directional ordering for school leaders (or for teachers, workplace leaders, even for employee samples more broadly). Given the dearth of research on the drivers of leader well-being (e.g., Barling & Cloutier, 2017; Kaluza et al., 2019), our study fills an important gap in the literature. We also identified critical methodological limitations in the design of longitudinal studies and appropriate statistical models to address these issues. From this perspective, our study is a substantive-methodological synergy. We demonstrated and extended state-of-the-art methodological research tools, addressing substantively essential matters with implications for theory, practice, and policy. Although our participants were school leaders, the methodological issues and extensions to CLPMs and RI-CLPMs should are likely to have broad cross-disciplinary relevance.

Directional Ordering: Limitations in Existing Research

For us, the most surprising limitation in this research literature was the lack of strong, longitudinal studies of the directional ordering of job satisfaction and self-efficacy in educational or organizational psychology. In many educational psychology studies, including leadership research, the relation is treated as a unidirectional directional ordering from self-efficacy (as the predictor) to job satisfaction as the (outcome), or directional ordering is not even considered (see related discussion by Wu & Griffin, 2012). However, almost all this research is cross-sectional, providing no basis for testing directional ordering. Indeed, authors of many of these studies recognized this limitation in their studies and called for stronger longitudinal data designs and improved statistical models to address this issue (e.g., Judge et al., 2020; Salanova et al., 2006; 2010; Skaalvik & Skaalvik, 2019; Skaalvik, 2020a, b; Zee & Koomen, 2016). Theoretically, a unidirectional model of directional-ordering is contrary to Bandura's theoretical model of reciprocal causation (e.g., Bandura, 1977, 1989, 1997; Maddux, 2009; Schunk & Pajares, 2005) and widely cited models in applied psychology (e.g., Bakker et al., 2014; Hobfoll, 1989; Fredrickson, 2001, 2008) based in part on Bandura's theoretical work.

Substantively, we found small but statistically significant and highly consistent reciprocal effects between job satisfaction and self-efficacy. Methodologically, our results were consistent across CLPMs and RI-CLPMs, alternative sets of invariance constraints (longitudinal equilibrium, full-trend stationarity, partial stationarity), and extensions of these models to include lag-2 and covariates. Indeed, the fit indices (RMSEA, TLI, and CFI) were nearly identical across the models invoking different constraints for both the CLPMs and the RI-CLPMs—including full-stationarity tests that are rarely or never used in CLPMs and RI-CLPMs.

Appropriate Interpretations of Within-Person (RI-CLPM) and Undecomposed Between-Person (CLPM) Effects

In educational psychology research, there is much confusion about the appropriate conclusions based on CLPMs and RI-CLPMs. The critical difference between these two models is CLPM's between-person (undecomposed, single-level) approach and RI-CLPMs' within-person (decomposed, multi-level) approach. CLPMs are applicable for comparisons between individuals. However, Hamaker et al. (2015; Mulder & Hamaker, 2021) and others are concerned that CLPMs confound within- and between-person effects in interpreting support for directional-ordering. RI-CLPMs evaluate prospective relations between temporary deviations from the trait level in one construct and temporary deviations from the trait level in a second construct (i.e., a between-person covariance of the within-person deviations). RI-CLPMs' auto-regressive (within-person) factors reflect deviations from the person's trait rather than the individual differences as in the CLPM. In the CLPM, the auto-regressive stability paths represent the stability of rank-order differences. In the RI-CLPM, the stability paths represent within-person carry-over effects or inertia (Hamaker et al., 2015), or slowly changing autoregressive factors (Kenny & Zautra (2001). Positive within-person stability paths indicate that scores higher than the global trait level in one wave are likely to be associated with higher scores in the next wave. That is, they have a lasting effect on subsequent measurement points in addition to the stability associated with the global trait factors. In CLPMs the cross-lagged paths represent undecomposed between-person processes. However, for RI-CLPMs, the cross-lagged paths reflect within-person processes. Thus, CLPMs and RI-CLPMs address different questions and typically offer different interpretations of the same data.

To illustrate this difference, for example, we ask the very basic question: What is the correlation between leaders' job satisfaction and self-efficacy? Based on our basic CLPMs and RI-CLPMs (Table 4), the answer varies depending on the perspective: .477 from a purely decomposed between-person (RI-CLPM) perspective; .230 from a purely within-person (RI-CLPM) perspective; and .356 from an undecomposed between-person single-level (CLPM) perspective. The single-level CLPM undecomposed estimate is a weighted average of the RI-CLPMs' within- and decomposed between-person estimates. Although different, none of these estimates is wrong, and each provides a different perspective. Concerning our question that focuses on relations between variables, the single-level estimate (.356) is probably more appropriate for descriptive purposes as a population estimate of the relation between the two constructs (also see correlations correlations in Table 2, mean r = .379). However, for other research questions, the other estimates might be more appropriate. Because the CLPMs and RI-CLPMs address different questions, their results lead to different interpretations. For the appropriate interpretation of results from the two approaches, see the wording of appropriate research questions associated with each outlined earlier (Marsh, Pekrun et al., 2022; Orth, et al., 2021; also see earlier discussion).

Juxtaposition of Results Based On CLPMs and RI-CLPMs

Much recent research treats CLPMs and RI-CLPMs as incompatible, competing models. Indeed, RI-CLPM's current popularity has resulted in a zeitgeist such that some researchers suggest that it is always more appropriate than the CLPM (but see discussion by Asendorpf, 2021; Lüdtke & Robitzsch, 2021; 2022; Marsh, Pekrun et al., 2022; Orth et al., 2021). However, either could be appropriate depending on the research question and the appropriate interpretation of the results. Furthermore, both perspectives provide potentially useful information about the substantive issue of directional ordering. Like Voelkle et al. (2014), we argue against viewing within- and between-person analyses as antagonistic approaches. Rather researchers should explore the juxtaposition between the two approaches. Here we explore differences between the two models in more detail.

In our study, the cross-lagged paths used to evaluate the directional-ordering of job satisfaction and self-efficacy are extraordinarily consistent over variations and extensions of RI-CLPMs and CLPMs. Indeed, for both CLPMs and RI-CLPMs, even the sizes of the reciprocal effects are very similar across the tests of directional predominance. Hence, support for our a priori predictions is consistent with both single-level undecomposed between-person (CLPM) and decomposed within-person (RI-CLPM) perspectives. Indeed, juxtaposing results and issues faced by both models provides a more robust interpretation of the results than would evaluating either model in isolation of the other. Of course, this is what we would expect if there was little between-person variance in the RI-CLPM. However, in our study, this similarity occurs even though much of the variance in the RI-CLPM was at the between-person level.

Methodological Issues

Goodness-of-Fit and the Measurement Model

As exemplified in our study, it is essential to evaluate models' goodness-of-fit and the measurement models' invariance over time. Unless there is good support for at least configural invariance, applying either CLPMs or RI-CLPMs is questionable. Furthermore, unless there is support for metric invariance of the factor structure over time, then constraining critical autoregressive parameters to be invariant over time is inappropriate. Metric invariance is particularly critical for RI-CPMs that are based on temporal deviation scores. Although the failure of metric invariance does not invalidate using CLPMs for prediction purposes, it complicates the interpretation of results. Hence, CLPMs and RI-CLPMs should always start by evaluating the underlying measurement model and support for invariance over the multiple time waves.

Because our focus has been on fully-latent models that control measurement error, we have not focused on this issue for manifest CLPMs and RI-CLPMs. Unaccounted measurement error and correlations of uniquenesses that are not accounted for bias results of both CLPMs and RI-CLPMs. This is worrisome because these issues are highly likely in cross-lagged panel data. Although it is possible to control for measurement error if there are reasonable estimates of reliability, such adjustments for longitudinal data are complex (particularly concerning correlated uniquenesses). Thus, fully latent models based on multiple indicators of all constructs provide the best control for measurement error.

Many models of longitudinal data, including CLPMs and RI-CLPMs, are nested under the corresponding measurement model (MM2 described in Tables 1 and 2). Thus, this measurement model's fit constitutes an essential comparison to subsequent CLPMs and RI-CLPMs. If the fit of the measurement model is meaningfully better than either CLPMs or RI-CLPMs, there is need for further exploration -- whether the misfit of the CLPMs or RI-CLPMs is substantively important concerning the interpretation of the results. For example, we found that the CLPMs' fits (Models MB1 and MB2 in Table 3) were good compared to traditional goodness-of-fit guidelines, but were noticeably poorer than the corresponding measurement model (MM2 in Table 1). Furthermore, a detailed evaluation of the measurement model offers better understanding of the data and the relations among the factors. This recommendation implies that multiple indicators are needed to test the measurement model. In summary, CLPM and RI-CLPM studies should routinely use fully-latent models, and preliminary analyses should evaluate support for the measurement model.

Goodness-of-Fit and Choice of Models

The CLPM is nested under the corresponding RI-CLPM. Hence, the fit of the RI-CLPM will always be better than that of the CLPM (at least for indices that do not correct for greater parsimony of the CLPM). The only exception would be the apparently unlikely situation when all global trait factors in RI-CLPMs have zero variances (Hamaker et al., 2015).

Relatedly, Orth et al. (2021; also see Asendorpf, 2021) contend that CLPMs and RI-CLPMs address different issues. Hence, model selection should reflect which model is more appropriate to the aims of the study as well as goodness-of-fit. Although we fully agree that CLPMs and RI-CLPMs address different questions, we add a caveat about the relevance of goodness-of-fit. The better fit of the RI-CLPM (and the corresponding measurement model) suggests systematic variation or covariation is unexplained by the CLPM. When the fit of RI-CLPM is better than that of the CLPM, the difference is likely due to RI-CLPMs' global trait factors. Consistently with Lüdtke and Robitzsch's (2021) simulation research, we found that the fit of the CLPM was substantially improved (approximating RI-CLPM's fit) by including Lag-2 and Lag-3 stability paths between non-adjacent waves of data. We suspect that this will often be the case for CLPMs (e.g., Lüdtke & Robitzsch, 2021; Marsh, Pekrun et al., 2022; Marsh, Pekrun, et al., 2018; also see related discussion by Mulder & Hamaker, 2021; Hamaker et al., 2015). Even without comparing CLPMs and the RI-CLPMs, it is possible to compare the fit of the basic CLPM with that of the final measurement model.

However, when additional lagged paths are included, it is essential to determine whether their inclusion alters substantively important parameter estimates and support for a priori hypotheses–a sensitivity test, as illustrated here.

Biases Associated with Omitted Covariates

We considered gender, age, role, and the sector as covariates. Although substantively interesting (Table 2), here we focus on whether their omission biases the estimates. Cross-lagged paths that are most important in testing directional ordering were nearly unaffected by the inclusion/exclusion of covariates in either CLPMs or the RI-CLPMs. Nevertheless, there will always be additional covariates that have not been measured. These additional, unmeasured covariates might be truly time-invariant covariates, time-invariant covariates that have different effects for different waves (perhaps reflecting additional, unmeasured process variables that vary with time and interact with the time-invariant covariates), time-varying covariates that might be specific to a particular wave, or even auto-regressive covariates that change gradually or systematically over time. However, the characteristics of these different biases and their likelihood of occurring has been given insufficient attention in CLPMs and RI-CLPMs (see Asendorpf, 2021; Lüdtke & Robitzsch, 2021; Schuurman & Hamaker, 2019).

For the CLPMs, truly time-invariant covariates will typically have their strongest direct effect on the first data wave (there may be exceptions, e.g., effects of gender which can change from before to after the onset of puberty). Thus, most of the covariates' effects across the nine waves (Model MC1 in Table 3) could be explained in their effects on the first wave of data (Model MC3). However, compared to CLPMs, RI-CLPMs provide better control for truly time-invariant covariates. This occurs because time-invariant covariates' effects in RI-CLPMs are largely absorbed by the global trait factors. Hence, in our study, the stability and cross-lagged paths for RI-CLPMs were largely unaffected by the exclusion of covariates.

For both CLPMs and RI-CLPMs, unmeasured time-varying covariates are worrisome confounders. However, adding lag-2 paths offers a stronger control for time-varying covariates (Marsh, Pekrun, et al., 2018, 2022; also see Lüdtke & Robitzsch, 2021). Thus, even if a covariate specific to Wave-T affects variables at Wave-T+1, it is less likely to affect variables at Wave-T+2 after controlling the effects from Wave-T and Wave-T+1 (see VanderWeele et al., 2020). Nevertheless, there exist more complex models that are stronger in terms of protecting against the effects of unmeasured covariates and random effects of other parameters (e.g., Lüdtke & Robitzsch, 2022; Orth et al., 2021; Usami et al., 2019a; 2019b; Voelkle et al., 2014). However, experience suggests that these models often have estimation problems that undermine their usefulness in applied research.

Strengths, Limitations, and Directions for Further Research

Generalizability

A strength of our study is the large, broadly representative sample of Australian school leaders enhances the generalizability of the results. Furthermore, our study is an exemplar in having nine annual waves based on large, representative samples. In this respect, it compares favorably with other research evaluating relations between job satisfaction and self-efficacy (both among leaders and among employees more broadly) and studies of directional ordering more generally (also see ***SM7*** on the use of self-report measures to assess job satisfaction and self-efficacy). However, further research is needed to test our findings' generalizability in other countries, school systems, and occupational groups.

Our focus on principals has important implications for the school leaders and the schools they lead. The focus of our research is on school principals' own wellbeing. However, there is also a need for further longitudinal research relating school leaders' wellbeing to school climate, teacher wellbeing, and student learning, achievement, and wellbeing.

Invariance Constraints

We note the largely idiosyncratic use of invariance constraints in previous applications of CLPMs and RI-CLPMs. Here we introduced different invariance constraints designed to address central issues. The first was the traditional set of factorial invariance constraints applied to the fully-latent, longitudinal measurement model (configural, strong, and strict invariance over time). Without a well-defined measurement model, the interpretation of structural models is questionable. Nevertheless, measurement models are often not considered, particularly in studies that rely on manifest CLPMs and RI-CLMS. Furthermore, the measurement model's fit affords a useful comparison for subsequent CLPMs and RI-CLPMs and a better understanding of the data (see Table 2).

The second was the set of longitudinal structural invariance constraints applied in the CLPMs and RI-CLPMs related to tests of stationarity (consistency of the estimates over time). The invariance constraints (longitudinal equilibrium and partial stationarity) reflect a blend of typical (e.g., longitudinal equilibrium), fully comprehensive (full-trend stationarity), and an expedient compromise (partial stationarity) that is easier to implement (see ***SM3- SM5*** for further discussion of stationarity). Although stationarity is a critical issue in many longitudinal models (e.g., Voelkle et al., 2014), it has been largely ignored, misunderstood, or misused in CLPMs and RI-CLPMs. This failure to test full-trend stationarity stems partly from the complex, non-linear invariance constraints required to test the invariance of latent variances. For this reason, we document its application using syntax that can be easily applied in other studies (see ***SM5***). Although there are clear statistical and substantive advantages to demonstrating full-trend stability, we leave as a question for further research whether tests of full-trend stationarity should be routinely incorporated in CLPMs and RI-CLPMs. We also note that it might be helpful to provide separate tests of multiple parameters constrained to be invariant within any one of our broader sets of constraints, particularly if tests of the overall constraint fail. Thus, it might be helpful to test the stability paths and cross-lagged paths in the longitudinal equilibrium constraint separately. Similarly, it might be helpful to test the invariance of residual variances and residual covariances separately. We also note that combining the measurement and structural invariance tests would be possible using the logic of MTMM analyses (see ***SM4***).

Finally, we re-introduce the notion of directional predominance–whether the cross-lagged paths leading from self-efficacy to job satisfaction are significantly larger than those leading from job satisfaction to self-efficacy. Although this issue was an early focus in cross-lagged panel studies (see Marsh, 1990), it has been largely ignored in CLPMs and RI-CLPMs. Here, we posed an interesting research question about the directional predominance of self-efficacy effects over job satisfaction effects and extended our models to test this issue. However, formal tests showed that the small positive cross-lagged paths leading from self-efficacy to job satisfaction did not differ significantly from cross-lagged paths leading from job satisfaction to self-efficacy. Tests of directional predominance may be useful in CLPMs and RI-CLPMs more generally.

Methodological Implications

Our research has implications for current research, and for the theoretical and statistical models that drive this research. In applied research, the link between self-efficacy and job satisfaction is typically seen as unidirectional (self-efficacy →job satisfaction). This unidirectional model suggests that interventions that enhance self-efficacy will improve job satisfaction. Nevertheless, we note that support for even this unidirectional model based on cross-sectional data is weak. Furthermore, because self-efficacy is typically very stable, the effectiveness of such interventions might be limited. Critically, this unidirectional perspective assumes that interventions to enhance job satisfaction will have no effect on self-efficacy. However, the demonstration of reciprocal relations between job satisfaction and self-efficacy fundamentally alters the suggestions based on this unidirectional perspective. Thus, the significant link from job satisfaction to self-efficacy implies that interventions that change job satisfaction will also influence self-efficacy. More importantly, interventions that target both job satisfaction and self-efficacy are likely to be particularly effective in enhancing both constructs. This expectation is consistent with the positive gain spirals hypothesized in several theoretical models (e.g., Conservation of Resources Theory, Hobfoll, 1989; broaden-and-build theory, Fredrickson, 2001, 2008; the job-demand resources model, Bakker et al., 2014; 2017; and social cognitive theory, Bandura, 1977).

We juxtapose and extend the major statistical models (CLPMs and RI-CLPMs) to test directional ordering. We recommend that researchers consider both CLPMs and RI-CLPMs, and draw conclusions from different models relevant to their research questions. However, even for studies pursuing causal inference from a potential outcome framework, Lüdtke and Robitzsch (2021) demonstrated circumstances in which estimates are biased for RI-CLPMs but not CLPMs and vice-versa. Nevertheless, they argued for a selection-on-observables approach in CLPMs, consistent with VanderWeele et al.'s (2019, 2020) perspective on causal inference with longitudinal data. Hence guidelines about the appropriate application and interpretation of these models are the basis of ongoing controversy; educational psychology and applied researchers need to understand the differences between the competing issues. The substantive issues addressed here are pertinent to educational and applied psychology. However, theoretical, design, interpretational, and statistical issues considered here will likely have broad cross-disciplinary generalizability.

Conclusions and Practical Implications

Barling and Cloutier (2017) argue that despite the considerable focus on employees' well-being, the well-being of school leaders has largely escaped attention. Barling and Cloutier emphasized the emotional toll of leadership and the failure of organizations to address their leaders' well-being (see also Kaluza et al., 2019; Skaalvik, 2020a, b). Supporting school leaders' job satisfaction and self-efficacy beliefs is vital for retaining motivated and capable leaders and attracting new leaders (Barling & Cloutier, 2017; also see Dicke, Marsh, et al., 2020). These challenges are particularly relevant for leaders in occupations with high levels of emotional labor and limited resources, such as educational leaders (Horwood et al., 2021; Riley et al., 2021). As highlighted by Riley et al. (2021; also see Horwood et al., 2021), Australian school leaders have faced steadily increasing demands and associated levels of stress with no actual increase in support services over the last decade.

School principals are in charge of large, multifaceted organizations and are under increasing public scrutiny from diverse stakeholders. School principals play a crucial role in our society, but report increasingly high demands, burnout, and attrition. Thus, policymakers must act to reverse this imminent crisis to ensure our schools and society flourish rather than flounder. Nevertheless, large-scale, longitudinal studies of principals' well-being are largely absent from occupational health, education, leadership, work, organizational, and psychology journals (but see Dicke et al., 2020; Dicke et al., 2022).

We evaluated the reciprocal effects linking school leaders' primary personal resource (their self-efficacy) and the most widely studied job-related well-being outcome (job satisfaction) for a large representative sample of Australian school principals. We positioned our research within the current controversy on how to evaluate reciprocal effects with cross-lagged panel designs. Finally, relating our study to new and evolving issues in evaluating models of reciprocal effects, we offered limitations of our research and directions for further research in this essentially unstudied field of the well-being of school principals.

In support of our a priori predictions and theoretical models, we found that school leaders' self-efficacy and job satisfaction are reciprocally related over an extended period of time. The reciprocal relations between self-efficacy and job satisfaction suggest an ongoing mutual intensification resulting in positive or negative feedback loops (Bakker & Demerouti, 2017; but see also Bandura, 1977). Positive gain spirals can be initiated by the positive effects of either self-efficacy or job satisfaction. Thus, successful interventions and strategies that increase school leaders' self-efficacy will likely enhance their job satisfaction. Likewise, interventions and strategies that improve job satisfaction are also likely to enhance self-efficacy. However, negative feedback loops can be triggered by adverse events that negatively influence either self-efficacy or job satisfaction. Thus, interventions should strive to simultaneously enhance self-efficacy, job satisfaction, and the reciprocal links between the two.

School principals need credible proactive feedback about their mental health and wellbeing. It is also essential that participants are confident about the confidentiality of the individual responses to the survey like that which is the basis of our study, one that is administered by a University Research Institute independent of State Departments of Education and other regulatory bodies. A critical feature of our ongoing research is to provide school principals immediate (within minutes of completing the survey) ongoing feedback that allows them to monitor critical mental health constructs compared to those experienced by other school principals and to changes in their own levels over time. This feedback includes "red flag" warnings suggesting that principals should consider seeking professional assistance. School principals view this feedback as the most valuable contribution of our research. However, the state and national school principal associations also value the aggregate reports on their members that helps them form policy within their organization. Additionally, results from our research have resulted in multiple Australian State Departments of Education policy changes to support the ongoing wellbeing of school principals. Hence school principals, professional organizations, and state and national regulatory and governmental agencies all value the annual reports from our survey that provide a collective voice for concerns faced by school principals.

We also emphasize the need for educational psychology research to achieve a constructive balance between methodological and substantive orientations. Support for this claim can be traced back to the 2007 special issue of Contemporary Educational Psychology devoted to the application of latent variable models in educational psychology and the need for substantive-methodological synergies. In the lead article of this special issue, Marsh and Hau (2007, also see SM8) first coined the term substantive-methodological synergy and presented a manifesto for a “substantive methodological synergy” movement. They recognized challenges in this approach, noting substantive-methodological synergies are criticized as either being too substantive (by methodological journals) or too methodological oriented (by substantive journals). From this perspective, the set of studies in the 2007 special issue of Contemporary Educational Psychology provided a welcome model, demonstrating how a variety of new latent variable models can be applied to substantively meaningful research issues based on data like that typically encountered in applied educational psychology research. We hope the present investigation contributes to bridging this gap in the tradition of Marsh and Hau (2007).

References

Asendorpf, J. B. (2021). Modeling developmental processes. In J. R. Rauthmann (Ed.), *Handbook of personality dynamics and processes* (pp. 815–835). London, UK. doi.org/10.1016/B978-0-12-813995-0.00031-5

Bakker, A B., & Demerouti, E. (2014). Job demands–resources theory. Well-being, 3, 1–28. In P. Chen & C. Cooper (Eds.), *Work and wellbeing: Wellbeing: A complete reference guide,* Volume III (pp. 1–28). Wiley. doi.org/10.1002/9781118539415.wbwell019

Bakker, A. B., Demerouti, E. (2017). Job demands–resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology, 22*(3), 273-285. doi.org/10.1037/ocp0000056

Bandura, A. (1977). Self-efficacy : Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. http://dx.doi.org/10.1037/0033-295X.84.2.191

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

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist*, *44*(9), 1175–1184. http://dx.doi.org/10.1037/0003-066X.44.9.1175

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

Bandura, A. (2008). Toward an agentic theory of the self. In H. Marsh, R. G. Craven, & D. M. McInerney (Eds.), Advances in Self Research, Vol. 3: Self-processes, learning, and enabling human potential (pp. 15-49). Charlotte, NC: Information Age Publishing.

Bandura, A., & Wood, R. (1989). Effect of perceived controllability and performance standards on self-regulation of complex decision making. *Journal of personality and social psychology*, *56*(5), 805–814. doi.org/1.1037/0022-3514.56.5.805.

Barling, J., & Cloutier, A. (2017). Leaders' mental health at work: Empirical, methodological, and policy directions. *Journal of Occupational Health Psychology, 22*(3), 394–406. doi.org/10.1037/ocp0000055

Berry, C. M., Carpenter, N. C., & Barratt, C. L. (2012). Do other-reports of counterproductive work behavior provide an incremental contribution over self-reports? A meta-analytic comparison. *Journal of Applied Psychology*, *97*, 613–636. doi.org/1.1037/a0026739

Berry, D., & Willoughby, M.T. (2017). On the practical interpretability of cross-lagged panel models: Rethinking a developmental workhorse. *Child Development*, *88*, 1186-1206.

Biaconcini, S., & Bollen, K.A. (2018). The latent variable-autoregressive laten trajectory model: A general framework for longitudinal data analysis. *Structural Equation Modeling*, *25*, 791-808,

Brunswick, E. (1952). The conceptual framework of psychology. Oxford, England: University Chicago Press.

Calsyn, R. J., & Kenny, D. A. (1977). Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievement? *Journal of Educational Psychology, 69*(2), 136.

Cattell, R. B. (1966). Patterns of change: Measurement in relation to state dimension, trait change, lability, and process concepts. *Handbook of Multivariate Experimental Psychology*, 355–402.

Chen, G., Gully, S.M. & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods, 4*(1), 62-83.

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*, 583-619.

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*, 879-894.

Darmody, M., & Smyth, E. (2016). Primary school principals' job satisfaction and occupational stress. *International Journal of Educational Management, 30,* 115–128. http://dx.doi.org/10.1108/IJEM-12-2014-0162

Dicke, T., Marsh, H. W., Parker, P. D., Guo, J., Riley, P., & Waldeyer, J. (2020). Job Satisfaction of Teachers and Their Principals in Relation to Climate and Student Achievement. Journal of educational psychology, 112(5), 1061-1073. https://doi.org/10.1037/edu0000409

Dicke, T., Marsh, H.W., Riley, P., Parker, P.D., Guo, J., & Horwood, M. (2018) Validating the Copenhagen Psychosocial Questionnaire (COPSOQ-II) Using Set-ESEM: Identifying Psychosocial Risk Factors in a Sample of School Principals. *Frontiers in Psychology, 9,* 584. doi.org/1.3389/fpsyg.2018.00584

Dicke, T., Parker, P.D., Guo, J., Basarkod, G., Marsh, H., Deady, M., Harvey, S., Riley, P. (2021). Ubiquitous emotional exhaustion in school principals: Stable trait, enduring autoregressive trend, or occasion-specific state? *Journal of Educational Psychology*

Dicke, T., Stebner, F., Linninger, C., Kunter, M., & Leutner, D. (2018). A longitudinal study of teachers' occupational well-being: Applying the job demands-resources model. Journal of Occupational Health Psychology, 23(2), 262–277. doi.org/1.1037/ocp0000070

Dimmock, C. & Hattie, J. (1996). School principals' self-efficacy and its measurement in a context of restructuring. School Effectiveness and School Improvement, 7(1),62-75, DOI: 10.1080/0924345960070103

Dormann, C., Fay, D., Zapf, D., & Frese, M. (2006). A state-trait analysis of job satisfaction: On the effect of core self-evaluations. *Applied Psychology: An International Review, 55* (1), 27-51.

Duffin, E. (2020). Number of elementary and secondary schools in the United States in 2017/2018, by school type. Statista, retrieved from https://www.statista.com/statistics/238307/number-ofuselementary-and-secondary-schools/

Enders, C. K. (2010). Applied Missing Data Analysis (illustrated). Guilford Press.

Federici, & Skaalvik, E. M. (2012). Principal self-efficacy : relations with burnout, job satisfaction and motivation to quit. Social Psychology of Education., 15(3), 295–320. https://doi.org/10.1007/s11218-012-9183-5

Federici, R.A., Skaalvik, E.M. (2011). Principal self-efficacy and work engagement: assessing a Norwegian Principal Self-Efficacy Scale. . Social Psychology of Education ,14, 575–600 (2011). https://doi.org/10.1007/s11218-011-9160-4

Fredrickson, B.L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist, 56*, 218–226.

Fredrickson, B. L. (2008). Promoting positive affect. In M. Eid & R. J. Larsen (Eds.), *The science of subjective well-being* (pp. 449–468). Guilford Press.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549–576. http://dx.doi.org/10.1146/annurev.psych.58.110405.085530

Grissom, J. A. & Bartanen, B. (2019). Principal effectiveness and principal turnover. Education Finance and Policy 14 (3), 355-382. https://doi.org/10.1162/edfp_a_00256

Hackett, R. D. (1989). Work attitudes and employee absenteeism: A synthesis of the literature. *Journal of Occupational Psychology, 62,* 235– 248. doi.org/1.1111/j.2044-8325.1989.tb004 95.x

Hallinger, P, Murphy, J (1985) Assessing the instructional management behavior of principals. The Elementary School Journal 86(2): 217–247.

Hallinger, P. (2010). Leadership for learning: Lessons from 40 years of empirical research. Journal of Educational Administration, 49, 125–142. https ://doi.org/10.1108/09578 23111 11166 99.

Hallinger, P. (2011). Leadership for learning: Lessons from 40 years of empirical research. Journal of Educational Administration, 49, 125–142. http://dx.doi.org/10.1108/09578231111116699

Hamaker EL, Kuiper RM, Grasman RPPP. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116. doi.org/10.1037/a0038889

Hamaker, E. L. (2018, November, 2018). How to run a multiple indicator RI-CLPM with Mplus. Retrieved from www.statmodel.com/download/RI-CLPM%20 Hamaker%20input.pdf

Hannah, S. T., Avolio, B. J., Luthans, F., & Harms, P. D. (2008). Leadership efficacy : Review and future directions. The Leadership Quarterly, 19(6), 669–692.

Hobfoll, S. E. (1989). Conservation of resources: A new attempt at conceptualizing stress. *American Psychologist 44*, 513–524.

Hofmans, J., Morin, A. J. S., Breitsohl, H., Ceulemans, E., Chénard-Poirier, L. A., Driver, C. C., et al., (2021). The baby and the bathwater: On the need for substantive-methodological synergy in organizational research. Industrial and Organizational Psychology, 14(4), 497-504. https://doi.org/10.1017/iop.2021.111

Horwood, M., Marsh, H.W. Parker, P. D., Riley, P., Guo, J. & Dicke, T. (2021). Burning Passion, Burning out: The Passionate School Principal, Burnout, Job Satisfaction, and Extending the Dualistic Model of Passion. *Journal of Educational Psychology.*

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

Hubner, N., Wagner, W., Zitzmann, S., & Nagengast, B. (2022, January 14). How causal is a reciprocal effect? Contrasting traditional and new methods to investigate the reciprocal effects model of self-concept and achievement. https:// doi. org/ 10. 31234/ osf. io/ f3e8w

Imbens, G. W., & Rubin, D. B. (2015). Causal inference for statistics, social, and biomedical sciences.

Cambridge University Press. https:// doi. org/ 10. 1017/ CBO97 81139 025751

Jelicić, H., Phelps, E., & Lerner, R. M. (2009). Use of missing data methods in longitudinal studies: the persistence of bad practices in developmental psychology. *Developmental Psychology*, *45*(4), 1195–1199. doi.org/1.1037/a0015665

Jöreskog, K. G. (1979). Statistical estimation of structural models in longitudinal investigations. (J. R. Nesselroade & B. Baltes, eds.). Academic Press.

Judge, T. A, & Bono, J. E. (2001). Relationship of core self-evaluations traits—self-esteem, generalized self-efficacy, locus of control, and emotional stability—with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology.*, *86*(1), 80–92. doi.org/1.1037/0021-901.86.1.80

Judge, T. A., Locke, E. A., & Durham, C. C. (1997). The dispositional causes of job satisfaction: A core evaluations approach. Research in Organizational Behavior, 19, 151–188.

Judge, T. A., Weiss, H. M., Kammeyer-Mueller, J. D., & Hulin, C. L. (2017). Job attitudes, job satisfaction, and job affect: A century of continuity and of change. *Journal of Applied Psychology*, *102*(3), 356-374. doi.org/10.1037/apl0000181

Judge, T. A., Zhang, S. C., & Glerum, D. R. (2020). Job satisfaction. In V.I Sessa & N. A Bowling, Eds). Essentials of Job Attitudes and Other Workplace Psychological Constructs, 207-241. Routledge. doi.org/10.4324/9780429325755

Kaluza, A. J., Boer, D., Buengeler, C., & van Dick, R. (2019). Leadership behaviour and leader self-reported well-being: A review, integration and meta-analytic examination. *Work & Stress*, 1-23.

Kasalak, G., & Dagyar, M. (2020). The Relationship between Teacher Self-Efficacy and Teacher Job Satisfaction: A Meta-Analysis of the Teaching and Learning International Survey (TALIS). *Educational Sciences: Theory and Practice*, *20*(3), 16-33.

Kenny, D A., & Zautra, A. (2001). Trait–state models for longitudinal data. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change.* (pp. 243–263). American Psychological Association. doi.org/1.1037/10409-008

Kenny, D. A. (1975). Cross-lagged panel correlation: A test for spuriousness. *Psychological Bulletin*, *82*(6), 887–903. doi.org/10.1037/0033-2909.82.6.887

Klassen, R. M., Tze, V. M., Betts, S. M., & Gordon, K. A. (2011). Teacher efficacy research 1998–2009: Signs of progress or unfulfilled promise?. *Educational Psychology Review*, *23*(1), 21-43. doi.org/1.1007/s10648-010-9141-8

Leithwood, K. & D. Jantzi, D. (2008). Linking leadership to student learning: The contributions of leader efficacy . Educational administration quarterly, 44(4), 496-528.

Leithwood, K. & Louis, K.S. (2012). Linking Leadership to Student Learning. San Francisco: Jossey-Bass.

Liu Y, Bellibaş M. Ş. &, Gümüş S. (2021). The effect of instructional leadership and distributed leadership on teacher self-efficacy and job satisfaction: Mediating roles of supportive school culture and teacher collaboration. Educational Management Administration & Leadership, 49(3), 430-453. doi:10.1177/1741143220910438

Locke, E.A., McClear, K. and Knight, D. (1996). Self-Esteem and Work. *International Review of Industrial and* Organizational Psychology, 11, 1–32.

Lüdtke, O., & Robitzsch, A. (2021, July 29). A Critique of the Random Intercept Cross-Lagged Panel Model. PsyArXiv. doi.org/10.31234/osf.io/6f85c

Lüdtke, O., & Robitzsch, A. (2022). A comparison of different approaches for estimating cross-lagged effects from a causal inference perspective. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*, 888----907. DOI: 10.1080/10705511.2022.2065278

Luszczynska, A., Scholz, U., & Schwarzer, R. (2005) The General Self-Efficacy Scale: Multicultural Validation Studies, The Journal of Psychology, 139:5, 439-457, doi.org/10.3200/JRLP.139.5.439-457

Macdonald, S. & Maclntyre, P. (1997). The Generic Job Satisfaction Scale. *Employee Assistance Quarterly.*, *13*(2), 1–16. doi.org/10.1300/J022v13n02_01

Maddux, J. E. (2009). Self-efficacy : The power of believing you can. In S. J. Lopez & C. R. Snyder (Eds.), *Oxford library of psychology: Oxford handbook of positive psychology* (pp. 335–343). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780195187243.013.0031

Maddux, J. E., & Gosselin, J. T. (2012). *Self-efficacy*. The Guilford Press.

Marsh, H.W., 1990. The causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis. *Journal of Educational Psychology, 82*, 646-656.

Marsh, H. W., Balla, J. R. & Hau, K-T. (1996). Assessing goodness of fit: Is parsimony always desirable? *The Journal of Experimental Education*, 64 (4), 364-390. http://dx.doi.org/10.1080/00220973.1996.10806604

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. doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., Dicke, T., Riley, P., Parker, P. D., Guo, J., Basarkod, G., & Martin, A. J. (2022). School principals' mental health and well-being under threat: A longitudinal analysis of workplace demands, resources, burnout, and well-being. Applied Psychology: Health and Well-Being. https://doi.org/10.1111/aphw.12423

Marsh, H. W., & Hau, K-T. (2007). Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. Contemporary Educational Psychology, 32, 151-171. http://dx.doi.org/10.1016/j.cedpsych.2006.10.008

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler’s (1999) Findings. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(3), 320–341. doi.org/1.1207/s15328007sem1103_2

Marsh, H. W, Huppert, FA, Donald, JN, Horwood, MS, & Sahdra, BK. (2020). The well-being profile (WB-Pro): Creating a theoretically based multidimensional measure of well-being to advance theory, research, policy, and practice. *Psychological Assessment, 32*(3), 294-313. doi.org/10.1037/pas0000787

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U. & Nagengast, B. (2010). A new look at the big-five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22,* 471-491. doi.org/10.1037/a0019227

Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J. S., & Von Davier, M. (2013). Why Item Parcels Are (Almost) Never Appropriate: Two Wrongs Do Not Make a Right—Camouflaging Misspecification With Item Parcels in CFA Models. *Psychological Methods,18,* 257-284. doi.org/10.1037/a0032773

Marsh, H. W., Morin, A. J. S., & Parker, P. D. (2015). Physical self-concept changes in a selective sport high school: A longitudinal cohort-sequence analysis of the big-fish-little-pond effect. *Journal of Sport and Exercise Psychology, 37*(2), 150–163. http://doi.org/10.1123/jsep.2014-0224

Marsh, H.W., Morin, A.J.S., Parker, P., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10*, 85-110

Marsh, Pekrun, R., & Lüdtke, O. (2022). Directional Ordering of Self-Concept, School Grades, and Standardized Tests Over Five Years: New Tripartite Models Juxtaposing Within- and Between-Person Perspectives. Educational Psychology Review. https://doi.org/10.1007/s10648-022-09662-9

Marsh, H. W., Pekrun, R., Murayama, K., Arens, A. K., Parker, P. D., Guo, J., & Dicke, T. (2018). An Integrated Model of Academic Self-Concept Development: Academic Self-Concept, Grades, Test Scores, and Tracking Over 6 Years. *Developmental Psychology*. http://dx.doi.org/10.1037/dev0000393

Marsh, H. W., Pekrun, R., Parker, P., Murayama, K., Guo, J. & Dicke, T. (2019).The Murky Distinction Between Self-Concept and Self-Efficacy : Beware of Lurking Jingle-Jangle Fallacies. *Journal of Educational Psychology, 111,* 331-353. doi.org/10.1037/edu0000281.

Marsh, H. W., & Scalas, L. F. (2018). Individually weighted-average models: Testing a taxonomic SEM approach across different multidimensional/global constructs because the weights “don’t make no nevermind”. *Structural Equation Modeling, 25*(1), 137-159. doi.org/10.1080/10705511.2017.1370377.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58,* 525-543.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY: Routledge.

Mincu, C. L. (2015). The impact of personal resources on organizational attitudes: Job satisfaction and trust in organisation. *Procedia – Social and Behavioral Sciences, 187,* 685–689.

Morin, A. J. S., Boudrias, J.-S., Marsh, H. W., McInerney, D. M., Dagenais-Desmarais, V., Madore, I., & Litalien, D. (2017). Complementary Variable- and Person-Centered Approaches to the Dimensionality of Psychometric Constructs: Application to Psychological Wellbeing at Work. Journal of Business and Psychology., 32(4), 395–419. https://doi.org/10.1007/s10869-016-

9448-7

Mulder, J. D., and Hamaker, E. L. (2021). Three extensions of the random intercept cross-lagged panel model. Structural Equation Modelin g. 28, 638–648. doi: 10.1080/10705511.2020.1784738

Mund, M., & Nestler, S. (2019). Beyond the cross-lagged panel model: Next-generation statistical tools for analysing dependencies across the life course. *Advances in Life Course Research, 41*, 100249.

Muthén, L. K., & Muthén, B. O. (2008-19). *Mplus User's Guide..* (Version 8) [Computer manual]. Muthén & Muthén.

Nübling, M., Burr, H., Moncada, S., and Kristensen, T. S. (2014). COPSOQ International Network: co-operation for research and assessment of psychosocial factors at work. Public Health Forum 22, A1–A3. doi: 10.1016/j.phf.2013.12.019

Nunez-Regueiro, F., Juhel, J., Bressoux, P., & Nurra, C. (2021). Identifying reciprocities in school motivation research: A review of issues and solutions associated with cross-lagged effects models. Journal of Educational Psychology. Advance online publication. https:// doi. org/ 10. 1037/ edu00 00700

Ock J. (2020), How satisfied are you with your job? Estimating the reliability of scores on a single-item job satisfaction measure. *International Journal of Selection and Assessment, 28,* 297– 309. doi.org/1.1111/ijsa.12285

Orth, U., Clark, D. A., Donnellan, M B., & Robins, R. W. (2021). Testing prospective effects in longitudinal research: Comparing seven competing cross-lagged models. *Journal of Personality and Social Psychology*., *120*(4), 1013–1034. doi.org/10.1037/pspp0000358

Pajares F. Self-Efficacy Beliefs in Academic Settings. *Review of Educational Research*. 1996;66(4):543-578. doi:10.3102/00346543066004543

Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2014). Juxtaposing math selfefficacy and self-concept as predictors of long-term achievement outcomes. Educational Psychology, 34(1), 29–48. https:// doi. org/ 10. 1080/ 01443 410. 2013. 797339

Pearl, J., Glymour, M., & Jewell, N. P. (2016). Causal inference in statistics: A primer. John Wiley & Sons.

Pekrun, R., Marsh, H. W., Suessenbach, F., Frenzel, A. C., & Goetz, T. (2022). School grades and students' emotions: Longitudinal models of within-person reciprocal effects. Learning and Instruction. Advance online publication. https://doi.org/10.1016/j.learninstruc.2022.101626

Pejtersen, J. H., Bjorner, J. B., and Hasle, P. (2010a). Determining minimally important score differences in scales of the Copenhagen Psychosocial Questionnaire. Scand. J. Public Health 38, 33–41. doi: 10.1177/1403494809347024

Pejtersen, J. H., Kristensen, T. S., Borg, V., and Bjorner, J. B. (2010b). The second version of the Copenhagen Psychosocial Questionnaire. Scand. J. Public Health 38, 8–24. doi: 10.1177/1403494809349858

Riley, P., See, S-M., Marsh, H., & Dicke, T. (2021) The Australian Principal Occupational Health, Safety and Wellbeing Survey (IPPE Report). Sydney: Institute for Positive Psychology and Education, Australian Catholic University

Salanova, M., Bakker, A., & Llorens, S. (2006). Flow at Work: Evidence for an Upward Spiral of Personal and Organizational Resources. *Journal of Happiness Studies, 7,* 1–22.

Salanova, M., Schaufeli, W. B., Xanthopoulou, D., & Bakker, A. B. (2010). The gain spiral of resources and work engagement: Sustaining a positive worklife. In A. B. Bakker (Ed.) & M. P. Leiter, *Work engagement: A handbook of essential theory and research* (p. 118–131). Psychology Press.

Schleicher, D. J., Hansen, D., & Fox, K. E. (2011). In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology: Vol 3. Maintaining, expanding, and contracting the organization, APA Handbooks in Psychology* (pp. 137–189). American Psychological Association.

Schunk, D. H., & Pajares, F. (2005). Competence Perceptions and Academic Functioning. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 85–104). Guilford Publications

Schuurman, N. K., & Hamaker, E. L. (2019). Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological Methods, 24*(1), 70–1. doi.org/10.1037/met0000188

Schwarzer, R., & Hallum, S. (2008). Perceived teacher self-efficacy as a predictor of job stress and burnout: Mediation Analyses. *Applied Psychology, 57,* 152–171. http://dx.doi.org/10.1111/j.1464-0597.2008 .00359.x

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), Causal and control beliefsMeasures in health psychology: A user's portfolio (pp. 35e37). NFER-NELSON.

Schutte, N. (2013). The broaden and build process: Positive affect, ratio of positive to negative affect and general self-efficacy. The Journal of Positive Psychology, 9(1), 66-74. DOI: 10.1080/17439760.2013.841280

Sherer, M., Maddux, J. E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. W. (1982). The Self-Efficacy Scale: Construction and validation. *Psychological Reports, 51,* 663-67.

Skaalvik, C. (2020a). School principal self-efficacy for instructional leadership: relations with engagement, emotional exhaustion and motivation to quit. Social Psychology of Education, 23(2), 479-498. DOI: 10.1007/s11218-020-09544-4

Skaalvik, C. (2020b). Self-efficacy for instructional leadership: Relations with perceived job demands and job resources, emotional exhaustion, job satisfaction, and motivation to quit. Social Psychology of Education, 23(5), 1343-1366. DOI: 10.1007/s11218-020-09585-9

Skaalvik and Skaalvik (2019), Teacher self-efficacy and collective teacher efficacy : Relations with perceived job resources and job demands, feeling of belonging, and teacher engagement. *Creative Education, 10,* 1400-1424 http://www.scirp.org/journal/ce

Skaalvik, E. M, & Skaalvik, S. (2017). Motivated for teaching? Associations with school goal structure, teacher self-efficacy, job satisfaction and emotional exhaustion. *Teaching and Teacher Education*, *67*, 152–16. doi.org/1.1016/j.tate.2017.06.006

Sniehotta, F. F., Scholz, U., & Schwarzer, R. (2005). Bridging the intention-behaviour gap: Planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. *Psychology and Health.*

Spector, P. E. (1997). *Job satisfaction: Application, assessment, causes, and consequences*. Sage.

Swann, W. B., Jr., Chang-Schneider, C., & McClarty, K. L. (2007). Do people’s self-views matter? Self-concept and selfesteem in everyday life. *American Psychologist*, *62*, 84-94. doi.org/10.1037/0003-066X.62.2.84

Tett, R. P., & Meyer, J. P. (1993). Job satisfaction, organizational commitment, turnover intention, and turnover: Path analysis based on meta-analytic findings. *Personnel Psychology*, *46*, 259–293. doi.org/1.1111/j.1744-657.1993.tb008 74.x

Tran H, McCormick J and Nguyen TT (2018) The cost of replacing south Carolina high school principals. Management in Education32(3): 109–118.

Tschannen-Moran M. & Gareis, C.R. (2007). Cultivating Principals' Self-Efficacy : Supports that Matter. *Journal of School Leadership, 17*(1), 89-114. doi.org/1.1177/105268460701700104

Tschannen-Moran, & Gareis, C. R. (2004). Principals' sense of efficacy . Journal of Educational Administration., 42(5), 573–585. https://doi.org/10.1108/09578230410554070

Usami, S., Todo, N., & Murayama, K. (2019). Modeling reciprocal effects in medical research: Critical discussion on the current practices and potential alternative models. *PLoS ONE, 14,* e0209133. http://dx.doi .org/10.1371/journal.pone.0209133

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods, 24,* 637–657. 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), 111e133. http://dx.doi.org/10.1207/s15326985ep3902_3.

VanderWeele, T. J. (2019). Principles of confounder selection. *European Journal of Epidemiology*, *34*, 211–219.

VanderWeele, T. J., Mathur, M. B., & Chen, Y. (2020). Outcome-wide longitudinal designs for causal inference: A new template for empirical studies. Statistical Science, 35 (3), 437–466. doi.org/10.1214/19-STS728

Voelkle, M. C. Brose, A., Schmiedek, F. & Lindenberger, U. (2014). Toward a Unified Framework for the Study of Between-Person and Within-Person Structures: Building a Bridge Between Two Research Paradigms. *Multivariate Behavioral Research*, *49*(3), 193–213. doi.org/10.1080/00273171.2014.889593

Wood, R., & Bandura, A. (1989). Social cognitive theory of organizational management. *Academy of Management Review*, *14,* 361–384. doi.org/10.2307/258173

Wu, C. H., & Griffin, M. A. (2012). Longitudinal relationships between core self-evaluations and job satisfaction. *Journal of Applied Psychology*, *97*(2), 331.

Zee, M., & Koomen, H. M. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research*, *86*(4), 981-1015.

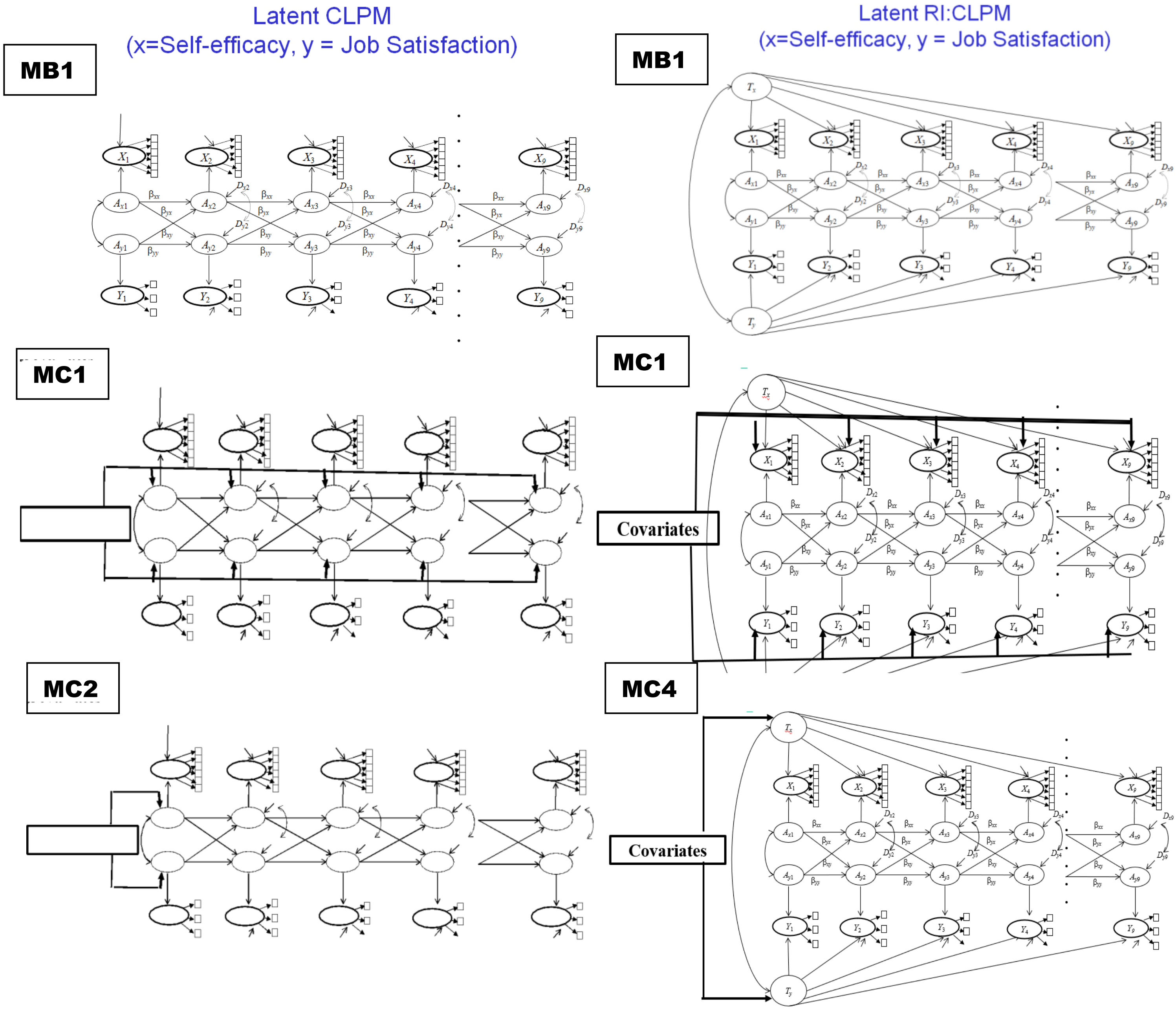
Zyphur, M.J., Allison, P.D., Tay, L., Voelkle, M.C., Preacher, K.J., Zhang, Z., Hamaker, E.L., Shamsollahi, A., Pierides, D.C., Koval, P., & Diener, E. (2020). From data to causes I: Building a general cross-lagged panel model. *Organizational research Methods*, *23*, 651-687

End Notes

1 We note that the fully latent RI-CLPM used here is like that originally presented by Hamaker (2018). However, we note that Mulder and Hamaker (2021) also additionally presented second fully latent RI-CLPM structure in which they started with having a random intercept for each indicator (the unlabelled boxes in Figure 1) and a within-wave factor model (see Figure 3 in Mulder & Hamaker, 2021). Interestingly, this alternative model was based in part on models of longitudinal models of longitudinal data developed by Marsh and Grayson (1994) in which they represented the longitudinal structure based on item factors (items loading on the same item from different waves and in Mulder and Hamaker's new model) or time-specific latent factors (as used here and originally proposed by Hamaker, 2018). Marsh and Grayson argued that in some circumstances these models are equivalent (e.g., when each factor is measure by three items and there are three waves). In general, however, these approaches will result in different estimates.

2 We test the statistical significance of the difference between the two cross-paths in the CLPMs and RL:CLPMs using two different approaches. First, we used the "model test" constraint in Mplus to test the difference between the two paths. Second, we actually fit a model in which the two cross-paths were constrained to be equal. In terms of the statistical significance of the difference between the two cross-paths, these models are equivalent. However, the second approach provided parameter estimates in which the two paths were the same. Interestingly this also resulted in substantially smaller standard errors.

Figure 1. Schematic Diagrams of Selected Cross-lag-panel models (CLPMs) and CLPMs with Random Intercepts (RI-CLPMs).



Note. MB = Basic Models; MC = Models with covariates, x = self-efficacy; y = job satisfaction. Each of these constructs is based on multi-item scales (the boxes). Not shown (to avoid clutter) are correlated uniquenesses relating responses to the same item across the multiple waves. For all measurement models there is scalar variance of estimates over time and over the multiple groups. In different models there are different sets of structural invariance over time and groups (e.g., invariance of stability- and cross-lagged paths). The primary difference between the CLPMs and RI-CLPMs is the inclusion of the Global Trait factors (Tx and Ty) representing the decomposed between-person component for each construct. Thus, in the CLPMs relations among the Ax and Ay factors in the CLPM represent relations among undecomposed between-person effects. In the RI-CLPMs these relations represent relations among within-person effects. In CLPMs with covariates shown here, covariates are posited to affect all 9 waves (MC1) or only the first wave (MC2). In RI-CLPMs with covariate models, covariates are posited to effect all (undecomposed) first-order factors (MC1) or only the latent trait factors. For all models shown here, only lag-1 paths are posited relating job satisfaction and self-efficacy in different waves (along with within-wave relations between the two constructs).

Table 1

Goodness of Fit for Confirmatory Factor Analysis (CFA) Measurement Model (MM): Invariance of the Measurement Factor Structure Over Multiple Waves 1 - 9

CFA Model	Chi-SQ	Df	RMSEA	CFI	TLI
CFA Models					
MM0 No invariance no correlated uniquenesses	19938	3826	.027	.851	.844
MM1 configural M0 with correlated uniquenesses	5613	3402	.011	.980	.976
MM2 metric M1 with factor loadings invariant	5762	3466	.011	.979	.975
MM3 Scalar M2 with intercepts invariant	6381	3546	.012	.974	.970

Note. Summary of goodness-of-fit statistics for different confirmatory factor analysis measurement model (mm): different factor analyses considered in the present investigation. Chi-SQ = chi-square; df = degrees of freedom; CFI = Comparative fit index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation. Model; INV = invariance constraints (constraining parameters to be invariant over time); CU = correlated uniqueness (relating residual variances associated with the same item over the multiple waves.

Table 2

Latent correlations between Self-efficacy (SE) and Job Satisfaction (JS) Over Nine Waves (2011-2019) and Background/Demographics

Factor	SE11	SE12	SE13	SE14	SE15	SE16	SE17	SE18	SE19	JS11	JS12	JS13	JS14	JS15	JS16	JS17	JS18	JS19
Self-Efficacy																		
SE11	**1.00**																	
SE12	**.68**	**1.00**																
SE13	**.68**	**.67**	**1.00**															
SE14	**.62**	**.67**	**.67**	**1.00**														
SE15	**.61**	**.62**	**.61**	**.68**	**1.00**													
SE16	**.57**	**.58**	**.60**	**.67**	**.68**	**1.00**												
SE17	**.53**	**.53**	**.58**	**.58**	**.61**	**.67**	**1.00**											
SE18	**.55**	**.53**	**.58**	**.59**	**.58**	**.65**	**.67**	**1.00**										
SE19	**.52**	**.54**	**.53**	**.54**	**.60**	**.60**	**.63**	**.72**	**1.00**									
Job Satisfaction																		
JS11	**.34**	**.33**	**.29**	**.31**	**.30**	**.25**	**.24**	**.24**	**.24**	**1.00**								
JS12	**.24**	**.38**	**.32**	**.35**	**.32**	**.28**	**.28**	**.29**	**.27**	**.64**	**1.00**							
JS13	**.27**	**.32**	**.39**	**.34**	**.33**	**.32**	**.30**	**.28**	**.31**	**.60**	**.70**	**1.00**						
JS14	**.23**	**.27**	**.30**	**.40**	**.34**	**.31**	**.28**	**.32**	**.32**	**.58**	**.63**	**.70**	**1.00**					
JS15	**.22**	**.25**	**.27**	**.34**	**.42**	**.31**	**.28**	**.33**	**.29**	**.52**	**.56**	**.61**	**.69**	**1.00**				
JS16	**.21**	**.26**	**.24**	**.30**	**.32**	**.36**	**.32**	**.32**	**.27**	**.48**	**.54**	**.56**	**.60**	**.67**	**1.00**			
JS17	**.15**	**.16**	**.24**	**.25**	**.28**	**.30**	**.39**	**.32**	**.28**	**.43**	**.51**	**.51**	**.57**	**.60**	**.65**	**1.00**		
JS18	**.17**	**.20**	**.27**	**.28**	**.25**	**.29**	**.32**	**.43**	**.33**	**.43**	**.44**	**.40**	**.50**	**.53**	**.55**	**.62**	**1.00**	
JS19	**.27**	**.29**	**.33**	**.30**	**.32**	**.30**	**.34**	**.39**	**.46**	**.55**	**.51**	**.54**	**.58**	**.60**	**.60**	**.61**	**.67**	**1.00**
Covariates																		
Gender	-.01	-.03	**-.06**	**-.07**	**-.06**	**-.05**	**-.04**	**-.05**	-.03	**-.05**	**-.04**	**-.05**	**-.06**	-.03	**-.04**	**-.05**	-.01	-.04
Age-linear	.02	**.07**	**.06**	.02	**.05**	.02	.02	.03	.02	**.10**	**.09**	**.08**	**.06**	**.09**	**.09**	**.14**	**.09**	**.14**
Age-Quad	.02	-.03	-.01	-.01	-.01	.00	.00	-.01	.00	-.01	.00	-.01	.03	-.01	.01	.00	-.01	.00
Position	**-.07**	**-.06**	**-.05**	**-.05**	**-.05**	**-.05**	-.04	-.04	-.04	**-.10**	**-.10**	**-.11**	**-.13**	**-.09**	**-.09**	**-.11**	**-.11**	**-.08**
Sector	**-.04**	-.01	-.04	-.01	-.02	-.04	-.01	.01	.01	**.17**	**.13**	**.11**	**.14**	**.12**	**.10**	**.09**	**.10**	**.13**

Note. Latent correlations between Self-efficacy (SE) and Job Satisfaction (JS) Over Nine Waves (2011-2019). Shaded correlations are between self-efficacy and job satisfaction within the same wave (*M r* = .316). Both constructs are very stable over time (*M lag 1 r = .679 for self-efficacy and .667 for Job satisfaction). Self-efficacy and job satisfaction in the same wave are moderately correlated (M r .379).* Correlations in bold are statistically significant (p < .05).

Table3

Relations between Self-efficacy (SE) and Job Satisfaction (JS) over time. Goodness of Fit Indices for Related Cross-Lag-Panel-Models (CLPMs) and Cross-Lag-Panel-Models with Random Intercepts (RI-CLPMs)

Model	CLPM					RI-CLPM				
Basic Models (MB)	Chi-SQ	Df	RMSEA	CFI	TLI	Chi-SQ	Df	RMSEA	CFI	TLI
MB1 no Invariance	7663	3578	.014	.962	.958	5939	3557	.011	.978	.975
MB2 Long Equility[a]	7737	3614	.014	.962	.958	6101	3618	.011	.977	.975
MB2 with CrsPath equality[a]	7737	3615	.014	.962	.958	6103	3619	.011	.977	.975
MB3. Part Stationarity	7753	3627	.014	.962	.958	6126	3624	.011	.977	.974
MB3 with CrsPath equality[a]	7753	3628	.014	.962	.958	6128	3625	.011	.977	.974
MB3. With Trait FL= free						6079	3608	.011	.977	.974
Lags, no Covariates (ML)										
ML1. Lag 2: CrsPath & Stability	6391	3604	.012	.974	.971	5996	3601	.011	.978	.975
ML2. Lag 2: Stability only	6397	3610	.012	.974	.971	6002	3607	.011	.978	.975
ML3. Lag 3: Stability only	6091	3593	.011	.977	.974	5955	3590	.011	.978	.976
Covariates, no lags (MC)										
MC1. COV: W1 to W9	8653	4059	.014	.958	.953	7084	4056	.011	.972	.969
MC2. COV: W1-W2	8854	4143	.014	.957	.952	7371	4131	.012	.97	.967
MC3. Cov: W1 only	8883	4155	.014	.957	.952	7400	4148	.012	.972	.969
MC4. COV: Paths fixed to zero	9105	4167	.014	.955	.95	7478	4164	.012	.97	.967
MC5. Cov: to Global Traits						7171	4152	.011	.972	.97
Lags + Covariates (MLC)										
MLC1. Lag-2 Cov-W1-W9	7341	4042	.012	.97	.966	6960	4039	.011	.973	.97
MLC2. Lag-2 Cov-W1-W2	7480	4132	.012	.969	.966	7209	4123	.012	.972	.969
MLC3. Lag-2 Cov-W1	7503	4144	.012	.969	.966	7252	4135	.012	.971	.969
MLC4. Lag-2 Cov-Fixed at 0	7749	4150	.012	.967	.964	7354	4147	.012	.971	.968
MLC5. Lag-2 Cov to Global Trait						7045	4135	.011	.973	.971
MLC6. Lag3-COV W1-W9	7046	4025	.012	.972	.969	6913	4022	.011	.973	.970
MLC7. Lag-3 Cov-Fixed at 0	7443	4133	.012	.907	.967	7307	4130	.012	.971	.968

Note. RMSEA = root mean square error of approximation, CFI = confirmatory fit index, TLI = Tucker-Lewis index, CrsPath = cross-lagged path; Stability = autoregressive stability paths; Cov = covariates, Lag 2 and Lag3= models with paths from each wave to the next two waves (Lag 2) or next three waves (Lag 3) in addition to Lag 1 paths. Models here extend the basic cross-lagged-panel models (CLPM) and the cross-lagged panel models with random intercepts (RI-CLPM; see Figure 1 and Table 4 for parameter estimates) by including additional lag-2 and lag-3 estimates (models ML1-Ml3), covariates (models MC1-MC5), and both additional lags and covariates (models MLC1-MLC7).

[a] We used the "model test" constraint in Mplus to test the difference between the two cross-lagged paths in theses CLPMs and RI-CLPMs. All six of these tests were non-significant (p > .10).

Table 4

Relations between Self-efficacy (SE) and Job Satisfaction (JS) over time. Parameter Estimates and Standard Errors (SEs) for Basic Models with Invariance Constraints Over the Multiple Waves (W1-W9)

Model	CLPM					RI-CLPM							
Basic Models (MB)	Cov SE-JS	Stab _JS	Stab _SE	CPth JStoSE	CPth SEtoJS	TrtVar JS	TrtVar SE	TrtCov SE-JS	Cov SE-JS	Stab _JS	Stab _SE	CPth JStoSE	CPth SEtoJS
MB2. Long Equilib	**.355**	**.685**	**.705**	**.065**	**.060**	**.535**	**.659**	**.477**	**.231**	**.270**	**.227**	**.063**	**.039**
SEs	*.023*	*.007*	*.007*	*.005*	*.005*	*.017*	*.018*	*.017*	*.042*	*.017*	*.018*	*.015*	*.016*
MB23 with CP equality	**.355**	**.684**	**.706**	**.063**	**.063**	**.536**	**.659**	**.477**	**.231**	**.262**	**.234**	**.053**	**.053**
SEs	*.023*	*.008*	*.007*	*.007*	*.008*	*.017*	*.018*	*.017*	*.042*	*.016*	*.017*	*.012*	*.012*
MB3. Part Stationarity	**.356**	**.679**	**.713**	**.059**	**.062**	**.531**	**.661**	**.477**	**.230**	**.280**	**.213**	**.066**	**.039**
SEs	*.023*	*.008*	*.008*	*.007*	*.008*	*.017*	*.017*	*.018*	*.042*	*.018*	*.018*	*.014*	*.017*
MB3 with CP equality	**.355**	**.680**	**.712**	**.060**	**.060**	**.533**	**.661**	**.476**	**.230**	**.271**	**.222**	**.056**	**.056**
SEs	*.023*	*.007*	*.007*	*.005*	*.005*	*.017*	*.017*	*.018*	*.042*	*.017*	*.017*	*.012*	*.012*
MB3 With Trait FL=free						**.496**	**.586**	**.482**	**.235**	**.274**	**.198**	**.063**	**.035**
SEs						*.028*	*.026*	*.017*	*.039*	*.018*	*.018*	*.014*	*.017*

Note. Basic cross-lag-panel models (CLPM) and CLPM with random intercept (RI-CLPM) were fit to the data (see Figure 1 and Table 3 for goodness-of-fit). Models 1-3 (MB1-MB3) refer to different sets of invariance constraints (MB2 = longitudinal equilibrium; MB3 = partial stationarity) over the multiple waves (W1-W9). For MB1 and MB2, an additional model was fit I which the two cross-products were constrained to be equal (CrsPath equality) to test for directional predominance. In all cases, the differences were non-significant (all ps > .10). Selected parameter estimates (see Figure 1) for the CLPM: within-wave correlation between self-efficacy and job satisfaction (Cov SE-JS) based on wave 1; the two autoregressive stability paths for SE and JS (Stab JS, Stab SE); and the two autoregressive cross-lagged paths from JS to SE (CPth JStoSE) and from SE to JS (CPth SEtoJS). Additional parameters for the RI-CLPM includes: decomposed between-person relative trait variance estimates JS and SE (VarBP JS, VarBP SE), and the within- and between-person covariances between SE and JS (BPCov_JS, WPCov_JS based on wave 1.

Table 5

Relations between Self-efficacy (SE) and Job Satisfaction (JS) over time. Extensions of Basic Cross-Lagged-Panel-Models (CLPMs) And Cross-Lagged-Panel-Models with Random Intercepts (RI-CLPMs) to Include Additional Lagged Autoregressive Paths and Covariates

	CLPM					RI-CLPM							
Models	Cov SE-JS	Stab _JS	Stab _SE	CPth JStoSE	CPth SEtoJS	TrtVar SEt	TrtVar JSt	TrtCov SE-JS	WPCov SE-JS	Stab _JS	Stab _SE	CPth JStoSE	CPth SEtoJS
Lags, No Covariates (ML)													
ML1. Lag 2: CrsPath & Stability	.34	.442	.429	.037	.045	.493	.636	.464	.218	.287	.271	.057	.067
SEs	*.023*	*.013*	*.012*	*.014*	*.016*	*.021*	*.022*	*.023*	*.048*	*.021*	*.020*	*.017*	*.020*
ML2. Lag 2: Stability only	.340	.442	.428	.042	.049	.497	.634	.482	.212	.288	.271	.053	.058
SEs	*.023*	*.013*	*.012*	*.008*	*.009*	*.020*	*.022*	*.020*	*.046*	*.021*	*.020*	*.017*	*.019*
ML3. Lag 3: Stability only	.337	.404	.383	.041	.048	.466	.617	.484	.210	.298	.302	.056	.072
SEs	*.023*	*.016*	*.015*	*.009*	*.010*	*.027*	*.029*	*.023*	*.048*	*.024*	*.020*	*.018*	*.020*
Covariates No lag (MC)													
MC1. COV: W1 to W9	.360	.663	.711	.059	.064	.515	.660	.477	.236	.284	.212	.067	.037
SEs	*.023*	*.008*	*.008*	*.007*	*.008*	*.017*	*.017*	*.018*	*.042*	*.018*	*.018*	*.014*	*.017*
MC2. COV: W1-W2	.359	.681	.714	.058	.061	.529	.667	.478	.224	.294	.221	.066	.054
SEs	*.023*	*.008*	*.008*	*.007*	*.008*	*.018*	*.018*	*.018*	*.044*	*.022*	*.021*	*.017*	*.020*
MC3. Cov: W1 only	.361	.684	.714	.059	.060	.533	.665	.480	.234	.300	.216	.060	.048
SEs	*.023*	*.008*	*.008*	*.007*	*.008*	*.018*	*.018*	*.018*	*.043*	*.019*	*.019*	*.015*	*.018*
MC4. COV: Paths fixed to zero	.356	.679	.713	.059	.062	.531	.661	.477	.229	.280	.213	.066	.039
SEs	*.023*	*.008*	*.008*	*.007*	*.008*	*.017*	*.017*	*.018*	*.042*	*.018*	*.018*	*.014*	*.017*
MC5. Cov to BP WP on Trait						.514	.658	.479	.228	.283	.213	.066	.037
SEs						.017	.017	.018	.042	.018	.018	.014	.017
Lags + Covariates (MLC)													
MLC1. Lag-2 Cov-W1-W9	.345	.44	.427	.042	.05	.480	.634	.483	.221	.294	.271	.055	.057
SEs	*.023*	*.013*	*.012*	*.008*	*.009*	*.021*	*.022*	*.021*	*.046*	*.021*	*.020*	*.017*	*.019*
MLC2. Lag-2 Cov-W1-W2	.346	.44	.428	.042	.048	.414	.641	.502	.198	.379	.269	.072	.039
SEs	*.023*	*.013*	*.012*	*.008*	*.009*	*.035*	*.023*	*.027*	*.048*	*.019*	*.018*	*.014*	*.018*
MLC3. Lag-2 Cov-W1	.347	.441	.428	.042	.048	.418	.634	.497	.219	.357	.275	.069	.047
SEs	*.023*	*.013*	*.012*	*.008*	*.009*	*.038*	*.023*	*.027*	*.046*	*.025*	*.020*	*.016*	*.019*
MLC4. Lag-2; Cov-Fixed at 0	.340	.442	.428	.042	.049	.497	.634	.482	.212	.288	.271	.053	.058
SEs	*.023*	*.013*	*.012*	*.008*	*.009*	*.020*	*.022*	*.02*	*.046*	*.021*	*.020*	*.017*	*.019*
MLC5. Lag2-COV Trait						.479	.631	.486	.212	.293	.272	.053	.056
SEs						*.021*	*.022*	*.021*	*.046*	*.021*	*.020*	*.017*	*.019*
MLC6. Lag-3 Cov-W1-W9	.342	.404	.383	.041	.048	.451	.619	.485	.218	.304	.302	.057	.070
SEs	*.023*	*.016*	*.015*	*.01*	*.010*	*.028*	*.029*	*.024*	*.048*	*.024*	*.020*	*.018*	*.020*
MLC7. Lag-3; Cov-Fixed at 0	.337	.404	.383	.041	.048	.466	.617	.484	.210	.298	.302	.056	.072
SEs	*.023*	*.016*	*.015*	*.009*	*.010*	*.027*	*.029*	*.023*	*.048*	*.024*	*.020*	*.018*	*.020*

Note. Models are extensions of Model MB3 (Partial stationarity, Tables 3 & 4) that include additional lagged estimates (ML) covariates (Cov, MC) or both MLC). W1-W9 = waves 1 to 9; Lag 2 and Lag3= models with paths from each wave to the next two waves (Lag 2) or next three waves (Lag 3) in addition to Lag 1 paths. CrsPath = autoregressive cross-lagged path; Stability = autoregressive stability paths. TrtVar and TrtCov are variance and covariances of the global trait factors,

Supplemental Materials

Section 1.
Bandura's conceptualization of self-efficacy and use of the term in organizational research.

Section 2.
Distinguishing Between RI-CLPM (Within-Person) and CLPM (Between-Person) Perspectives

Section 3.
Stationarity, Measurement, and Structural Invariance Constraints.
Derivation and for Constraints to Test Full-trend stationarity

Section 4.
Full Stationarity in relation to the MTMM matrix

Section 5.
Mplus Syntax for implementing Full-trend stationarity

Section 6.
Missing Data in Relation to our Sampling Design

Section 7.
The Reliance on Self-report Measures.

Section 8. The Need for Substantive-Methodological Synergy in Contemporary Educational Psychology

Section 1
Bandura's conceptualization of self-efficacy and use of the term in organizational research.

The self-efficacy label is broadly used in different ways in psychological research and research more generally. Bandura focused primarily on task-specific measures of self-efficacy that conformed to specific guidelines (e.g., Bandura, 2006) designed to eliminate most of the evaluative component of self-perceptions (see discussion by Marsh et al., 2018, on the Murky distinction between self-efficacy and self-concept). According to Bandura's guidelines, task-specific self-efficacy items define a specific task and the standards against which individuals judge themselves. For example, the self-efficacy item "I can run 100 meters in 13 seconds in the next school track meet" clearly specifies the task and the standard (13 seconds). Importantly, the response is purely descriptive and not evaluative (i.e., whether this would represent a great result—a personal best, or a terrible one—the slowest time this season). Similarly, in an educational context, Betz and Hackett’s (1983) math self-efficacy scale asks, “How confident are you to be able to work out the price of a t-shirt when getting 20% off?” Thus, items that do not clearly present both a specific task and the standards against which they are to be evaluated might be useful in evaluating self-beliefs (e.g., self-concepts), but do not correspond to the design features originally proposed by Bandura and other self-efficacy researchers (see Marsh et al., 2018 for further discussion). Self-efficacy measures typically used in organizational psychology research do not satisfy these criteria—particularly the explicit inclusion in the self-efficacy items of objective standards the respondents use to make judgments—but instead focus on more generalized perceptions of self-efficacy in workplaces and fail to provide an objective standard of what constitutes success. It might be more appropriate to refer to these measures as domain-specific measures of self-belief or self-concept measures rather than task-specific (particularly in relation to Bandura's guidelines).

We suggest that the distinction between task and generalized self-efficacy is really a continuum rather than a dichotomy. Purely task-specific items – particularly those that conform to Bandura's (2006) guidelines -- are likely to be so highly contextualized that they are largely idiosyncratic to a particular study. These self-efficacy items might be closely related to future outcomes that are also highly contextualized and idiosyncratic to the particular study. However, these measures probably not be useful in other settings. In this respect, it would be necessary to develop new self-efficacy measures for each study. Klassen et al. (2011) similarly noted that Self-efficacy measures are most predictive of future behaviors when both the self-efficacy and outcome measures are narrowly defined. However, the self-efficacy measures lose generalizability to other settings as specificity increases. Similarly, Pajares (1996) warned that “microscopically operationalized” measures lose practical utility, even as predictive power increases. Hence, for broad, generic constructs like job satisfaction, a generalized measure of self-efficacy is likely to be more predictive than an overly specific task measure of self-efficacy. Schwarzer and Hallum (2008) found that a task-specific and their Generalized Self-efficacy scale (like what we used in the present investigation) were highly correlated and similarly related to job stress and burnout. In a longitudinal analysis, they found that task-specific and generalized measures of self-efficacy measures defined a single self-efficacy latent construct.

We also note that the specificity matching principle (e.g., Brunswick, 1952; Swan, et al. 2007) suggests that the breadth and generality of a predictor variable should match the outcome that it is designed to predict. Thus, in relation to high task-specific outcomes, it is appropriate to use highly specific self-efficacy items. However, it is more appropriate to use more generalized self-efficacy measures for more broadly generalizable outcomes such as job satisfaction.

Our measure of self-efficacy is a generalized measure that falls closer to the generalized end of this continuum than a task-specific measure. Nevertheless, even though the respondents were not asked to respond specifically in relation to the workplace, the items were part of an extensive survey specific to the workplace. Thus, in the preamble to the survey, respondents were told: *The survey is being conducted in response to concerns that the increasing complexity and workload demands of school leadership roles are impacting on the health and wellbeing of Australian school leaders.* Hence, respondents responded to these items concerning their work context. In this respect, the measure was more like a domain-specific measure than a purely generalized measure that was domain-general.

We note that there is an ambiguity in organizational psychology (and more generally) in the conceptualization of self-efficacy – in relation to Bandura's (2006) criteria of item design and the relevance to task-specific, domain-specific, and generalized measures of self-efficacy. The resolution of this issue is beyond the scope of the present investigation. Nevertheless, it is clearly justifiable to test the generalizability of predictions based on Bandura's conceptual model in relation to generalized measures of self-efficacy.

Marsh, et al. (2018) extensively evaluated the murky distinction between self-concept and self-efficacy in relation to Bandura's (2006) guidelines for the appropriate construction of self-efficacy items. The study was a substantive-methodological synergy, bringing together new substantive, theoretical and statistical models, and developing new tests of the classic jingle-jangle fallacy. They noted the many possible differences between self-concept and self-efficacy, but argued that the most critical was the appropriate construction of self-efficacy items in relation to Bandura's guidelines. They argued and demonstrated empirically that scales that were claimed to measure self-efficacy but did not incorporate the standard against which responded judged their success were indistinguishable from self-concept scales, but were distinct from true self-efficacy measures that were consistent with Bandura's principles for the design of self-efficacy items. The critical distinction is that self-concept items are heavily influenced by the evaluation of success on the items that invoke social comparison processes. Because self-efficacy items were judged in relation to an objective standard (e.g., whether I can run 100 meters in 13 seconds) the social comparison processes were largely eliminated.

Marsh, et al. (2018) demonstrated that in a representative sample of 3,350 students from math classes in 43 German schools, generalized math self-efficacy and math outcome expectancies were indistinguishable from math self-concept, but were distinct from two math self-efficacy scales that consisted of items consistent with Bandura's (2006 design principles.

In big-fish-little-pond effect school-average achievement has a negative effect on math self-concept, even though individual math achievement has a positive effect. The finding is one of the most cross-nationally generalizable findings in psychological research. The big-fish-little-pond effect based substantially on social comparison processes. Marsh et al. (2018) demonstrated that, consistent with a priori predictions that the big-fish-little-pond was strong for measures of generalized math self-efficacy, math outcome expectancies, and math self-concept but were completely eliminated in responses to the two true self-efficacy measures. This is consistent with Bandura's argument that appropriately constructed self-efficacy items largely eliminate the evaluative (social comparison) component that is so strong in self-concept and related self-belief measures. Hence the critical issue is to supply an objective standard against which to evaluate success that largely removes the social comparison processes. From this perspective, the critical difference between self-efficacy and other self-belief constructs is the inclusion of an objective standard rather than where the self-belief construct is classified in relation to task specificity, domain specificity, or a generalized construct.

After controlling for pre-test variables (including prior achievement), each of the three self-concept-like constructs (math self-concept, outcome expectancy, and generalized math self-efficacy) in each of the four years of secondary school was more strongly related to post-test outcomes (school grades, test scores, future aspirations) than were the corresponding two true self-efficacy -like factors.

Extending discussion by Marsh et al. (1997), Marsh et al. (2019) clarified distinctions between self-efficacy and self-concept; the role of evaluation, worthiness, and outcome expectancy in self-efficacy measures; and complications in generalized and global measures of self-efficacy. This prior research is relevant to the present investigation and organizational research more generally because there are apparently no measures of self-efficacy research routinely used in organizational psychology research that satisfy Bandura's (2006) guidelines. Hence, we argue that these measures should be labelled as self-concept (or some related self-belief term) rather than self-efficacy.

Section 2. Distinguishing Between RI-CLPM (Within-Person) and CLPM (Between-Person) Perspectives

Historically, CLPMs are the most widely used model in psychological research to test directional ordering based on panel data. However, following Hamaker and colleagues (Hamaker, 2018; Hamaker et al., 2015; Hamaker & Muthén, 2020; Mulder & Hamaker, 2021), the popularity of RI-CLPMs has surged. However, Orth et al. (2021) stressed that these two models address different questions, result in different interpretations, and make different assumptions. Here we argue that both models separately and their juxtaposition are relevant to different theoretical questions. Hence understanding these differences is critical.

Structural Differences

The primary difference between the RI-CLPM and the CLPMs is structural. In particular, the RI-CLPM includes a stable trait factor (Tx and Ty in Figure 1), but the CLPM does not. Hence, CLPMs are nested under the RI-CLPM. In this sense, RI-CLPMs model how scores obtained at each wave differs from a person's stable trait (a within-person perspective disaggregated from the between-person perspective reflected in the stable trait). Furthermore, these time-specific within-person variations observed at one point in time influence within-person-variations observed at later time points. In contrast, CLPMs model how inter-individual differences observed at one point in time are related to inter-individual differences observed at later time points (a between-person perspective).

The critical difference in distinguishing between these models is in how the term between-person is used. In the CLPM, between-person effects reflect a combination of deviations from a global trait (within-person effects) and a stable trait (between-person effects). This use of the term between-person effects is consistent with its use in most studies of individual differences in relations among variables and in most cross-sectional studies. In contrast, the RI-CLPM decomposes these effects into separate within- and between-person components. The use of the generic term between-person is appropriate within the context of each of these two models. However, to highlight this distinction and avoid confusion, we use the terms “decomposed between-person effects” (RI-CLPM) and “undecomposed between-person effects” (CLPM).

For both CLPMs and RI-CLPMs, the critical parameters are stability paths (Bxx and Byy in Figure 1), and particularly the cross-lagged paths (Bxy and Byx in Figure 1). If both Bxy and Byx are statistically significant, X and Y are said to be reciprocally related. If both these cross-lagged paths are statistically significant but differ significantly in size, one of the constructs is directionally predominant. For example, if paths from prior self-efficacy to subsequent job satisfaction were greater than paths from prior job satisfaction to self-efficacy, self-efficacy would be said to be directionally predominant in job satisfaction.

Most applications, particularly the RI-CLPM and comparisons of the two models, are based on manifest variables. However, stronger versions of CLPMs and RI-CLPMs are possible when based on latent variables in which multiple indicators are used to define each construct (Marsh, et al., 2022; Mulder & Hamaker, 2021; also see Hamaker, 2018).

CLPMs and RI-CLPMs: Distinct Research Questions

In longitudinal panel studies, the multiple time waves (level 1) are nested under the individual person (level 2). The level 2 variables are each principal's unweighted averages of job satisfaction and self-efficacy over time (i.e., the random intercepts). These level 2 effects represent stable traits that are consistent for a given school principal over time but differ from principal to principal. The CLPM is analogous to a single-level model, evaluating relations between job satisfaction and self-efficacy (within-waves and over time) without controlling for person-level differences in these variables. Thus, the CLPM’s undecomposed between-person effect does not separate within-person and between-person components. In contrast the RI-CLPM takes a within-person perspective, estimating relations between job satisfaction and self-efficacy after controlling between-person stable trait effects (person-level intercepts). These two perspectives are easily confused. However, they address distinct research questions and often result in different interpretations. To clarify this distinction, for each model, we offer the following research questions (also see Marsh, Pekrun et al., 2022; Orth et al., 2021):

- **CLPMs:** When leaders have high self-efficacy (relative to other leaders), do they experience a subsequent rank-order change in job satisfaction (relative to other leaders)? Likewise, when leaders have high job satisfaction (relative to other leaders), do they experience a subsequent

rank-order change in self-efficacy (relative to other leaders)? Thus, do individual differences in self-efficacy positively predict rank-order change in relative job satisfaction, and do individual differences in job satisfaction positively predict rank-order change in relative self-efficacy?
- **RI-CLPMs:** When school leaders experience higher than their usual self-efficacy (relative to their long-term average self-efficacy), do they experience a subsequent change in job satisfaction (relative to their long-term average job satisfaction)? Likewise, when leaders experience higher than their usual job satisfaction (relative to their long-term average job satisfaction), do they experience a subsequent change in self-efficacy (relative to their long-term average self-efficacy?

The potential confounding of effects with unmeasured covariates threatens the interpretation of both CLPMs and RI-CLPMs. The RI-CLPM's key advantage is providing increased control for time-invariant (between-person) covariates not included (e.g., Hamaker et al., 2015; Mulder & Hamaker, 2021; Marsh, et al., 2022). The rationale for this claim is that true time-invariant (between-person) covariates mainly affect the global trait factors. These global train effects are statistically independent of the within-person autoregressive factors. Thus, unmeasured covariates that are truly time-invariant are most likely to influence the sizes of the (decomposed between-person) global trait factors, but are unlikely to affect the within-person autoregressive factors used to evaluate directional ordering. This distinction is important in interpreting the results, particularly tests of directional ordering and directional predominance. Of course, it is possible to include measured covariates into RI-CLPMs; Mulder and Hamaker (2021) described how to do this with manifest RI-CLPMs. Although CLPMs can also control measured covariates, they provide less control for the confounding effects of unmeasured covariates.

RI-CLPMs and CLPMs provide only weak control for unmeasured covariates that vary over time. Similarly, neither provides good control for time-invariant covariates whose effects vary from wave to wave, possibly reflecting an unmeasured process). Even here, however, the invariance of the effects over waves would suggest that these potentially confounding covariates do not substantially affect the results. Nevertheless, the best way to control the effects of unmeasured (time-varying and time-invariant) covariates is to include them in the study's design and incorporate them into the statistical models. Hence the selection of covariates is crucial for RI-CLPMs and particular CLPMs. Hence, it is disappointing that alternative strategies for the inclusions of covariates have been given so little attention to designing, analyzing, and interpreting CLPMs and RI-CLPMs (also see VanderWeele, 2019).

Juxtaposing CLPMs and RI-CLPMs: The Controversy

The underlying conceptual and statistical issues posed by CLPMs and RI-CLPMs are quite different and address fundamentally different questions. The relative appropriateness of between- and within-person empirical paradigms generally, the CLPM and the RI-CLPM specifically, and the appropriate interpretations of findings from each are currently hot substantive and methodological topics. Although between-person perspectives dominate educational psychology research, there have been increasing calls for the use of within-person and related person-centered approaches. CLPMs (a between-person perspective) are usually used to test unidirectional, bidirectional, and reciprocal effects. However, there has been a dramatic increase in the use of RI-CLPMs (a within-person perspective), heated debates about their relative usefulness, and unwarranted claims about the relevance of each.

Researchers (e.g., Berry & Willoughby, 2017; Hamaker et al., 2015; Mund & Nestler, 2019) have emphasized the CLPM's inability to separate within (state-like) and between (trait-like) effects. These concerns have led to diverse models specifically designed to address this limitation (e.g., Biaconcini & Bollen, 2018; Curran et al., 2014; Hamaker et al., 2015; Mund & Nestler, 2019; Zyphur et al., 2020). This tsunami in RI-CLPM's popularity has resulted in a zeitgeist among some applied researchers, suggesting that the RI-CLPM is always more appropriate than the CLPM.

In contrast to this surge in RI-CLPMs' popularity, several researchers (Asendorpf, 2021; Hubner et al., 2022; Lüdtke & Robitzsch, 2021; Marsh, et al., 2022) critiqued the RI-CLPM from the perspectives of the mathematical derivation, causal inference from a potential outcome framework (e.g., Imbens & Rubin, 2015; Pearl et al., 2016), and simulated data. Similar to our study, Lüdtke and Robitzsch's overarching aim was "to provide a more balanced discussion of two main approaches (CLPM and RI-CLPM) for analyzing cross-lagged panel designs, and we would like to emphasize that—despite recent methodological recommendations—there are still good reasons to use the traditional CLPM when estimating cross-lagged effects" (p. 3). They found that RI-CLPMs' estimates of cross-lag effects were often biased in different data-generating scenarios with fixed confounders (e.g., demographic variables) with time-varying effects. Following Marsh, Pekrun et al. (2018; also see Marsh, Pekrun et al., 2022), Lüdtke and Robitzsch emphasized the inclusion lag-2 effects to provide more robust controls for confounding (VanderWeele et al., 2020). Furthermore, the goodness-of-fit of CLPMs with lag-2 effects variables was similar to the fit of RI-CLPMs. Thus, goodness-of-fit is no longer a critical issue in the comparison of RI-CLPMs and CLPMs with lag-2 effects. Furthermore, Lüdtke and Robitzsch argued for a selection-on-observables approach to the control for covariates using information observed in the data (previous measures of the outcomes and additional covariates) used in CLPMs rather than the inclusion of stable trait factors in RI-CLPMs. VanderWeele et al. (2019; 2020) similarly argue for a selection-on-observables approach to causal inference with longitudinal data. Following Lüdtke and Robitzsch's recommendations, we extended traditional CLPMs based on lag-1 estimates to include lag-2 estimates.

In summary, RI- CLPMs and CLPMs address distinct questions and typically have different or even contradictory interpretations. Their comparative strengths, weaknesses, and appropriate interpretations are the basis of ongoing controversy and debate. Hence, educational and psychological researchers need to understand these differences in CLPMs and RI-CLPMs. Our study is the first to juxtapose their theoretical rationale and results in tests of the directional ordering of job satisfaction and self-efficacy. We aim this presentation to educational psychology researchers based on a classic issue that is surprisingly understudied. Further, we demonstrate extensions of RI-CLMSs, establish the benefits of lag-2 effects, and show their usefulness in educational and psychological research.

Section 3. Stationarity, Measurement, and Structural Invariance Constraints.

Because we test fully latent models, we can explore invariance issues not previously considered with CLPMs and RI-CLMPs. In pursuit of this extension, we discuss the invariance of the measurement model that has been widely considered in CLPMs and for simulated data in the Mulder and Hamaker (2021) latent RI-CLPMs. We extend this discussion to include longitudinal invariance, including formal tests of stationarity that apparently have not been applied to either CLPMs or RI-CLPMs. In each step of this process, the overarching questions are to what extent there is support for the generalizability of interpretations (particularly Research Hypothesis 1) concerning the invariance constraints and the implications of failure to support invariance. We discuss these issues in detail because these issues are not well understood and are largely unconsidered in studies of the relations between job satisfaction and self-efficacy.

A Well-Defined Measurement Model.

When the multiple indicators of each construct are parallel over the different time waves, testing the invariance of the measurement structure over time is critical. In fully-latent models of longitudinal data, it is typical to evaluate a set of models with different invariance constraints of parameters over time (e.g., Marsh, Morin et al., 2014; Meredith, 1993; Millsap, 2011; Mulder & Hamaker, 2021): configural, with no invariance constraints; metric, with factor loading invariance; and scalar, with intercept invariance. Following Marsh et al. (2013; for further discussion, see Marsh et al., 1996; Joreskog, 1979), we also tested correlated uniquenesses relating residual variance terms for the same item in different waves. Failure to include them typically results in biased parameter estimates and a poorly fitting model.

For CLPMs and RI-CLPMs, the most critical tests are for configural and metric invariance. Intercept invariance is an important characteristic in some longitudinal models, but not for CLPMs and RI-CLPMs based on covariance structures. These initial models test whether the measurement model is well-defined. These tests are not based on any particular model (e.g., CLPM or RI-CLPM). However, applying either CLPMs or RI-CLPMs models is dubious unless there is reasonable support for at least configural invariance. Furthermore, if support for the invariance of factor loadings is weak, then tests of invariance tests for other associated parameters (e.g., stability and cross-lagged paths in CLPMs and RI-CLPMs) are questionable. This is a critical issue in that many cross-lag panel studies are based on manifest models that preclude tests of the measurement model.

Longitudinal Structural Invariance Constraints. Particularly when there is good support for metric invariance, this structure should be the starting point for fully-latent CLPMs and RI-CLPMs. When there are three or more waves of data, there are practical, conceptual, theoretical, and methodological reasons for testing the invariance (i.e., equivalence) of parameters over time. From a practical perspective, as the number of waves increases, invariance over time greatly enhances model parsimony and facilitates the summary of the results. Thus, researchers can present one set of results rather than separate results for each wave.

Conceptually, the invariance of parameters over time suggests that the results may generalize over the time frame considered. Theoretically, particularly in developmental studies, there may be a theoretical basis for positing that parameters vary over time (e.g., there are particular onset effects; constructs become more or less distinct with age and maturation). Statistically, the imposition of invariance constraints increases the power of critical tests. However, because there is no clear consistency in invariance constraints used in applied research, we briefly review alternative strategies based on historical and current practice.

For CLPMs and RI-CLPMs, the main focus is on the cross-lagged paths (Bxy and Byx in Figure 1) used to test directional ordering but also the autoregressive stability paths (Byy and Bxx in Figure 1). However, the other auto-regressive parameters (e.g., residual variances and covariances in waves 2-9, and the variances and covariances in wave 1) are also relevant. A logical approach would be to test these parameters sequentially (e.g., Morin et al., 2017). For present purposes, we test three a priori constraints that we apply to our CLPMs and RI-CLPMs.

- **Longitudinal Equilibrium**: For the CLPMs with at least three waves of data (see Figure 1), Marsh, Pekrun, et al. (2018; also see Parker et al., 2014) used the term "development equilibrium" for tests of invariance over time of stability (Byy, Bxx,) and cross-paths (Bxy, Byx). Their focus was on how these paths varied with age for school students. Nevertheless, these tests are also considered in typical in RI-CLPMs that do not have a specific developmental focus (e.g., Mulder & Hamaker, 2021). Support for these invariance constraints facilitate the interpretation of results. Hence, to broaden the use of this terminology, we use the term longitudinal equilibrium (also see Morin et al., 2017, who used the term predictive equilibrium).

- **Partial Stationarity.** In tests of partial stationarity, in addition to constraints in Longitudinal Equilibrium, we also impose the invariance of all residual variances following the first wave (i.e., waves 2-9 in the present investigation). We also constrained the residual within-wave covariances between job satisfaction and self-efficacy to be invariant over waves 2-9. In each case, wave 1 was excluded because the corresponding parameter estimate in wave 1 is a covariance (rather than a residual covariance) and a variance (rather than a residual variance). Because variance and covariance estimates for wave 1 are not incorporated into the invariance constraints in this model, we refer to it as partial stationarity rather than full-trend stationarity.

Full-trend stationarity. Theoretical and empirical studies of longitudinal data (e.g., Voelkle et al., 2014) emphasize the importance of full-stationarity in facilitating the interpretation and robustness of interpretations. By definition, full stationarity includes the invariance of variance estimates over time. Although early cross-lag panel studies (e.g., Kenny, 1975) emphasized stationarity, this feature has not been incorporated into current CLPMs and RI-CLPMs. It is straightforward to test the invariance of residual variances and covariances for Ax and Ay factors for waves 2-9 (Figure 1)—what we refer to as partial stationarity. However, the corresponding estimates for the first wave are variances and covariances rather than residual variances and covariances. Hence, it is not straightforward to test the invariance of variances and covariances across all waves. Nevertheless, in a different context, Kenny and Zautra (2001) demonstrated complex non-linear constraints to test the invariance of variances and covariances across all waves, including wave 1. However, because CLPMs and RI-CLPMs are models of the covariance

Implementing Stationarity Constraints: Derivation and for Constraints to Test Full-trend stationarity

In the following, we briefly explain the nonlinear constraints that need to be specified in order to implement stationarity for the autoregressive process in the CLPM or RI-CLPM. Let us assume a bivariate first-order autoregressive process with two variables A_{xt} and A_{yt} (t = 2,…,T):

$$A_{xt} = \beta_{xx}A_{x,t-1} + \beta_{xy}A_{y,t-1} + D_{xt}$$

$$A_{yt} = \beta_{yy}A_{y,t-1} + \beta_{yx}A_{x,t-1} + D_{yt}$$

where the coefficients $®_{xx}$ and $®_{yy}$ represent the (time-invariant) autoregressive effects from $A_{x,t-1}$ to A_{xt}, and from $A_{y,t-1}$ to A_{yt}. The coefficients $®_{xy}$ and $®_{yx}$ represent the (time-invariant) cross-lagged effects from $A_{y,t-1}$ to A_{xt}, and from $A_{x,t-1}$ to A_{yt}. The disturbance terms D_{xt} and D_{yt} denote random components that are unrelated to previous time points and are normally distributed with zero means, variances ϕ_{Dx} and ϕ_{Dy}, and covariance ϕ_{Dxy}. Note that with multiple indicators A_{xt} and A_{yt} represent the latent factors in the CLPM or the autoregressive traits in the RI-CLPM.

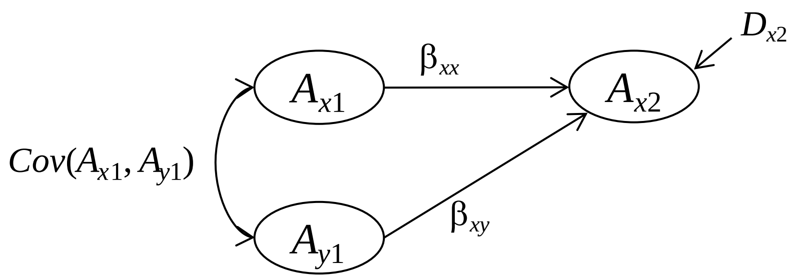
Under the assumption of stationarity, the variances (ϕ_{Ax} and ϕ_{Ay}), and covariance (ϕ_{Axy}) of the factors A_{xt} and A_{yt} are assumed to be constant across time. In order to implement stationarity in SEM software, three nonlinear constraints for the variances and covariances of the disturbances need to be specified (Kenny & Zautra, 1995). First, applying the formula for R^2 in a regression with two predictors, the variance of the disturbance needs to be constrained as follows:

$$\phi_{Dx} = \phi_{Ax} - \beta_{xx}^2\phi_{Ax} - \beta_{xy}^2\phi_{Ay} - 2\beta_{xx}\beta_{xy}\phi_{Axy}$$

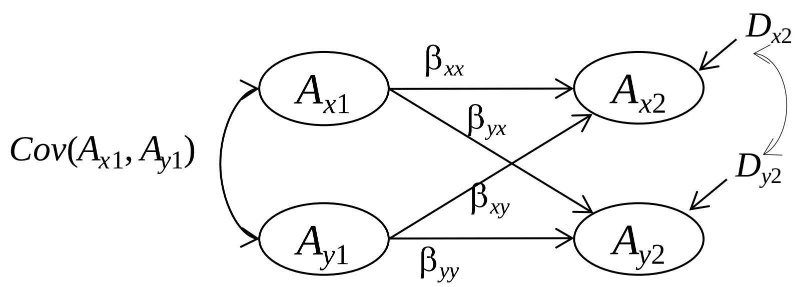
A similar constraint needs to be specified for ϕ_{Dy}^2. Furthermore, the constraint for the covariance of the disturbances is given by:

$$\phi_{Dxy} = \phi_{Axy}(1 - \beta_{xx}\beta_{yy} - \beta_{xy}\beta_{yx}) - \beta_{xx}\beta_{yx}\phi_{Ax} - \beta_{yy}\beta_{xy}\phi_{Ay}.$$

Variance Constraint



Covariance Constraint



Section 4. Full Stationarity in relation to the MTMM matrix

We treated the set of constraints on the measurement model and structural model separately in the present investigation. However, following Kenny (1975), full stationarity might be evaluated with constraints on the MTMM matrix, extending constraints on the measurement model to include variances and covariances. In particular, because all CLPMs and RI-CLPMs are nested under the final measurement model, support to full-stationarity based on the measurement model would demonstrate support for full-stationarity based on subsequent models. This would have the advantage of simplifying some of the constraints (e.g., invariance of variances would no longer require complex non-linear constraints). Also, tests of these constraints would be preliminary and not tied to any particular model, and may even inform analysts of the most appropriate models to consider. More broadly, considering a diverse set of invariance constraints demonstrates their usefulness in CLPMs and RI-CLPMs more generally. This is particularly relevant because many applications consider only a single set of invariance constraints without evaluating their appropriateness or fully understanding their implications.

Section 5. Mplus Syntax for implementing Full-trend stationarity

```
 USEVARIABLES ARE
 JS1_2019 JS2_2019 JS3_2019 JS4_2019
 SE1_2019 SE2_2019 SE3_2019 SE4_2019 SE5_2019 SE6_2019
 JS1_2018 JS2_2018 JS3_2018 JS4_2018
 SE1_2018 SE2_2018 SE3_2018 SE4_2018 SE5_2018 SE6_2018
 JS1_2017 JS2_2017 JS3_2017 JS4_2017
 SE1_2017 SE2_2017 SE3_2017 SE4_2017 SE5_2017 SE6_2017
 JS1_2016 JS2_2016 JS3_2016 JS4_2016
 SE1_2016 SE2_2016 SE3_2016 SE4_2016 SE5_2016 SE6_2016
 JS1_2015 JS2_2015 JS3_2015 JS4_2015
 SE1_2015 SE2_2015 SE3_2015 SE4_2015 SE5_2015 SE6_2015
JS1_2014 JS2_2014 JS3_2014 JS4_2014
SE1_2014 SE2_2014 SE3_2014 SE4_2014 SE5_2014 SE6_2014
JS1_2013 JS2_2013 JS3_2013 JS4_2013
SE1_2013 SE2_2013 SE3_2013 SE4_2013 SE5_2013 SE6_2013
JS1_2012 JS2_2012 JS3_2012 JS4_2012
SE1_2012 SE2_2012 SE3_2012 SE4_2012 SE5_2012 SE6_2012
JS1_2011 JS2_2011 JS3_2011 JS4_2011
SE1_2011 SE2_2011 SE3_2011 SE4_2011 SE5_2011 SE6_2011
 ;

MISSING ARE ALL (-99);

IDVARIABLE = ID2019;

ANALYSIS:  ESTIMATOR=MLR;coverage = .05;

 model:
!Factor loadings -- invariant over time

!!!! Translate factors into X and Y !!Let  x Job Satisfac5tion &  Y  = Self-efficacy ;
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
X1 by  JS1_2011@0.743   JS2_2011 JS3_2011 JS4_2011                   (fl1-fl4);
X2 by  JS1_2012@0.743   JS2_2012 JS3_2012 JS4_2012                   (fl1-fl4);
X3 by  JS1_2013@0.743   JS2_2013 JS3_2013 JS4_2013                   (fl1-fl4);
X4 by  JS1_2014@0.743   JS2_2014 JS3_2014 JS4_2014                   (fl1-fl4);
X5 by  JS1_2015@0.743   JS2_2015 JS3_2015 JS4_2015                   (fl1-fl4);
X6 by  JS1_2016@0.743  JS2_2016 JS3_2016 JS4_2016                    (fl1-fl4);
X7 by  JS1_2017@0.743   JS2_2017 JS3_2017 JS4_2017                   (fl1-fl4);
X8 by  JS1_2018@0.743   JS2_2018 JS3_2018 JS4_2018                   (fl1-fl4);
X9 by  JS1_2019@0.743   JS2_2019 JS3_2019 JS4_2019                   (fl1-fl4);

Y1 by  SE1_2011@0.694   SE2_2011 SE3_2011 SE4_2011 SE5_2011 SE6_2011 (f21-f26);
Y2 by  SE1_2012@0.694   SE2_2012 SE3_2012 SE4_2012 SE5_2012 SE6_2012 (f21-f26);
Y3 by  SE1_2013@0.694   SE2_2013 SE3_2013 SE4_2013 SE5_2013 SE6_2013 (f21-f26);
Y4 by  SE1_2014@0.694   SE2_2014 SE3_2014 SE4_2014 SE5_2014 SE6_2014 (f21-f26);
Y5 by  SE1_2015@0.694   SE2_2015 SE3_2015 SE4_2015 SE5_2015 SE6_2015 (f21-f26);
Y6 by  SE1_2016@0.694   SE2_2016 SE3_2016 SE4_2016 SE5_2016 SE6_2016 (f21-f26);
Y7 by  SE1_2017@0.694   SE2_2017 SE3_2017 SE4_2017 SE5_2017 SE6_2017 (f21-f26);
Y8 by  SE1_2018@0.694   SE2_2018 SE3_2018 SE4_2018 SE5_2018 SE6_2018 (f21-f26);
Y9 by  SE1_2019@0.694   SE2_2019 SE3_2019 SE4_2019 SE5_2019 SE6_2019 (f21-f26);

!!!!!!!!!!Correlated uniqueness -- invariant over time if keep constraints!!!!!!!
JS1_2011 JS1_2012 JS1_2013 JS1_2014 JS1_2015 JS1_2016 JS1_2017 JS1_2018 JS1_2019 with
JS1_2011 JS1_2012 JS1_2013 JS1_2014 JS1_2015 JS1_2016 JS1_2017 JS1_2018 JS1_2019;!(cu1);
JS2_2011 JS2_2012 JS2_2013 JS2_2014 JS2_2015 JS2_2016 JS2_2017 JS2_2018 JS2_2019 with
```

```
  JS2_2011 JS2_2012 JS2_2013 JS2_2014 JS2_2015 JS2_2016 JS2_2017 JS2_2018 JS2_2019;!(cu2);
  JS3_2011 JS3_2012 JS3_2013 JS3_2014 JS3_2015 JS3_2016 JS3_2017 JS3_2018 JS3_2019 with
  JS3_2011 JS3_2012 JS3_2013 JS3_2014 JS3_2015 JS3_2016 JS3_2017 JS3_2018 JS3_2019;!(cu3);
  JS4_2011 JS4_2012 JS4_2013 JS4_2014 JS4_2015 JS4_2016 JS4_2017 JS4_2018 JS4_2019 with
  JS4_2011 JS4_2012 JS4_2013 JS4_2014 JS4_2015 JS4_2016 JS4_2017 JS4_2018 JS4_2019;!(cu4);
  SE1_2011 SE1_2012 SE1_2013 SE1_2014 SE1_2015 SE1_2016 SE1_2017 SE1_2018 SE1_2019 with
  SE1_2011 SE1_2012 SE1_2013 SE1_2014 SE1_2015 SE1_2016 SE1_2017 SE1_2018 SE1_2019;!(cu5);
  SE2_2011 SE2_2012 SE2_2013 SE2_2014 SE2_2015 SE2_2016 SE2_2017 SE2_2018 SE2_2019 with
  SE2_2011 SE2_2012 SE2_2013 SE2_2014 SE2_2015 SE2_2016 SE2_2017 SE2_2018 SE2_2019;!(cu6);
  SE3_2011 SE3_2012 SE3_2013 SE3_2014 SE3_2015 SE3_2016 SE3_2017 SE3_2018 SE3_2019 with
  SE3_2011 SE3_2012 SE3_2013 SE3_2014 SE3_2015 SE3_2016 SE3_2017 SE3_2018 SE3_2019;!(cu7);
  SE4_2011 SE4_2012 SE4_2013 SE4_2014 SE4_2015 SE4_2016 SE4_2017 SE4_2018 SE4_2019 with
  SE4_2011 SE4_2012 SE4_2013 SE4_2014 SE4_2015 SE4_2016 SE4_2017 SE4_2018 SE4_2019;!(cu8);
  SE5_2011 SE5_2012 SE5_2013 SE5_2014 SE5_2015 SE5_2016 SE5_2017 SE5_2018 SE5_2019 with
  SE5_2011 SE5_2012 SE5_2013 SE5_2014 SE5_2015 SE5_2016 SE5_2017 SE5_2018 SE5_2019;!(cu9);
  SE6_2011 SE6_2012 SE6_2013 SE6_2014 SE6_2015 SE6_2016 SE6_2017 SE6_2018 SE6_2019 with
  SE6_2011 SE6_2012 SE6_2013 SE6_2014 SE6_2015 SE6_2016 SE6_2017 SE6_2018 SE6_2019;!(cu10);

  !!!!!!!!!!!RESID VARiances  -- invariant over time if keep constraints!!!!!!!!!!
  JS1_2011 JS1_2012 JS1_2013 JS1_2014 JS1_2015 JS1_2016 JS1_2017 JS1_2018 JS1_2019;!(RV1);
  JS2_2011 JS2_2012 JS2_2013 JS2_2014 JS2_2015 JS2_2016 JS2_2017 JS2_2018 JS2_2019;!(RV2);
  JS3_2011 JS3_2012 JS3_2013 JS3_2014 JS3_2015 JS3_2016 JS3_2017 JS3_2018 JS3_2019;!(RV3);
  JS4_2011 JS4_2012 JS4_2013 JS4_2014 JS4_2015 JS4_2016 JS4_2017 JS4_2018 JS4_2019;!(RV4);

  SE1_2011 SE1_2012 SE1_2013 SE1_2014 SE1_2015 SE1_2016 SE1_2017 SE1_2018 SE1_2019;!(RV5);
  SE2_2011 SE2_2012 SE2_2013 SE2_2014 SE2_2015 SE2_2016 SE2_2017 SE2_2018 SE2_2019;!(RV6);
  SE3_2011 SE3_2012 SE3_2013 SE3_2014 SE3_2015 SE3_2016 SE3_2017 SE3_2018 SE3_2019;!(RV7);
  SE4_2011 SE4_2012 SE4_2013 SE4_2014 SE4_2015 SE4_2016 SE4_2017 SE4_2018 SE4_2019;!(RV8);
  SE5_2011 SE5_2012 SE5_2013 SE5_2014 SE5_2015 SE5_2016 SE5_2017 SE5_2018 SE5_2019;!(RV9);
  SE6_2011 SE6_2012 SE6_2013 SE6_2014 SE6_2015 SE6_2016 SE6_2017 SE6_2018 SE6_2019;!(RV10);

  !!!!!!!!!!!!INTERCEPTS  -- invariant over time if keep constraints;
  [JS1_2011 JS1_2012 JS1_2013 JS1_2014 JS1_2015 JS1_2016 JS1_2017 JS1_2018
JS1_2019];!(IT1);
  [JS2_2011 JS2_2012 JS2_2013 JS2_2014 JS2_2015 JS2_2016 JS2_2017 JS2_2018
JS2_2019];!(IT2);
  [JS3_2011 JS3_2012 JS3_2013 JS3_2014 JS3_2015 JS3_2016 JS3_2017 JS3_2018
JS3_2019];!(IT3);
  [JS4_2011 JS4_2012 JS4_2013 JS4_2014 JS4_2015 JS4_2016 JS4_2017 JS4_2018
JS4_2019];!(IT4);

  [SE1_2011 SE1_2012 SE1_2013 SE1_2014 SE1_2015 SE1_2016 SE1_2017 SE1_2018
SE1_2019];!(IT5);
  [SE2_2011 SE2_2012 SE2_2013 SE2_2014 SE2_2015 SE2_2016 SE2_2017 SE2_2018
SE2_2019];!(IT6);
  [SE3_2011 SE3_2012 SE3_2013 SE3_2014 SE3_2015 SE3_2016 SE3_2017 SE3_2018
SE3_2019];!(IT7);
  [SE4_2011 SE4_2012 SE4_2013 SE4_2014 SE4_2015 SE4_2016 SE4_2017 SE4_2018
SE4_2019];!(IT8);
  [SE5_2011 SE5_2012 SE5_2013 SE5_2014 SE5_2015 SE5_2016 SE5_2017 SE5_2018
SE5_2019];!(IT9);
  [SE6_2011 SE6_2012 SE6_2013 SE6_2014 SE6_2015 SE6_2016 SE6_2017 SE6_2018
SE6_2019];!(IT10)

  !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
  !!!!!!!!!!  Transforming X (Job Satisfaction) into between-person variables for X  !!!!

  !!!!!!!!!!  creates a new set of latent ix factors (one for each wave);
  iX1 by X1;
  iX2 by X2;
  iX3 by X3;
```

```
iX4 by X4;
iX5 by X5;
iX6 by X6;
iX7 by X7;
iX8 by X8;
iX9 by X9;
!!!!!!!!!! iX factors fixed factor loadings = 1 on auto-regressive (AR) X factors
AR1_X by iX1@1;
AR2_X by iX2@1;
AR3_X by iX3@1;
AR4_X by iX4@1;
AR5_X by iX5@1;
AR6_X by iX6@1;
AR7_X by iX7@1;
AR8_X by iX8@1;
AR9_X by iX9@1;

!!!!Constrain x & xi (Job Satisfaction) factors to have zero variance & intercepts;
X1-X9@0; iX1-iX9@0;

!X factors fixed factor loadings = 1 on global X Trait factors;
trait_X by iX1-iX9@1;

trait_X (VGTR_X); ! variance of X trait factor;

!!!!  set of 8 stability paths constrained to be invariant over time
!!!!  b_XoX = path from AR_X Factor (Time = T) to AR_X factor (Time = T+1)--stability
  AR2_X on AR1_X (b_XoX);
  AR3_X on AR2_X (b_XoX);
  AR4_X on AR3_X (b_XoX);
  AR5_X on AR4_X (b_XoX);
  AR6_X on AR5_X (b_XoX);
  AR7_X on AR6_X (b_XoX);
  AR8_X on AR7_X (b_XoX);
  AR9_X on AR8_X (b_XoX);


AR1_X (VAR1_X);        ! Variance of AR_X for T1;
AR2_X-AR9_X (RVAR_X);  ! Resid Variance of AR_X for T2-T9 (not T1), invariant over time;


!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!!!!  Transforming Y (Self-Efficacy ) into bivariate Start variables for Y  !!!!

!!!!!!!!!!  creates a new set of latent ix factors (one for each wave);
iY1 by Y1;
iY2 by Y2;
iY3 by Y3;
iY4 by Y4;
iY5 by Y5;
iY6 by Y6;
iY7 by Y7;
iY8 by Y8;
iY9 by Y9;
!!!!!!!!!! iY factors fiYed factor loadings = 1 on auto-regressive (AR) Y factors
AR1_Y by iY1@1;
AR2_Y by iY2@1;
AR3_Y by iY3@1;
AR4_Y by iY4@1;
AR5_Y by iY5@1;
AR6_Y by iY6@1;
```

```
AR7_Y by iY7@1;
AR8_Y by iY8@1;
AR9_Y by iY9@1;

!!!!Constrain original Y & Yi (JS) factors to have zero variance & intercepts;
Y1-Y9@0; iY1-iY9@0;

!Y factors fixed factor loadings = 1 on global Y Trait factors;
trait_Y by iY1-iY9@1;

trait_Y (VGTR_Y); !VT = variance of Y trait factor;

!!!!  set of 8 stability paths constrained to be invariant over time;
!!!!  b_YoY = path from AR_Y Factor (Time = T1) to AR_Y factor (Time = T+1)--stability;
  AR2_Y on AR1_Y (b_YoY);
  AR3_Y on AR2_Y (b_YoY);
  AR4_Y on AR3_Y (b_YoY);
  AR5_Y on AR4_Y (b_YoY);
  AR6_Y on AR5_Y (b_YoY);
  AR7_Y on AR6_Y (b_YoY);
  AR8_Y on AR7_Y (b_YoY);
  AR9_Y on AR8_Y (b_YoY);

AR1_Y (VAR1_Y);        ! Variance of AR_Y for T1;
AR2_Y-AR9_Y (RVAR_Y);  ! Resid Variance of AR_Y for T2-T9 (not T1), invariant over time;

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!!!!  Relations Between X & Y  !!!!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

!!!!!!!  auto-regressive cross-lagged paths
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!  b_XoY = path from AR_Y Factor (Time = T) to AR_X factor (Time = T+1); cross paths
 AR2_X on AR1_Y (b_XoY);
 AR3_X on  AR2_Y (b_XoY);
 AR4_X on AR3_Y (b_XoY);
 AR5_X on  AR4_Y (b_XoY);
 AR6_X on  AR5_Y (b_XoY);
 AR7_X on  AR6_Y (b_XoY);
 AR8_X on  AR7_Y (b_XoY);
 AR9_X on  AR8_Y (b_XoY);


!!!!  b_YoX = path from AR_X Factor (Time = T) to AR_Y factor (Time = T+1); cross paths
 AR2_Y on AR1_X (b_YoX);
 AR3_Y on AR2_X (b_YoX);
 AR4_Y on  AR3_X (b_YoX);
 AR5_Y on AR4_X (b_YoX);
 AR6_Y on AR5_X (b_YoX);
 AR7_Y on AR6_X (b_YoX);
 AR8_Y on AR7_X (b_YoX);
 AR9_Y on AR8_X (b_YoX);


! covariances Between X & Y State (ST) factors;
!STate1_X with !STate1_Y (Cov!ST);
!STate2_X with !STate2_Y (Cov!ST);
!STate3_X with !STate3_Y (Cov!ST);
!STate4_X with !STate4_Y (Cov!ST);
!STate5_X with !STate5_Y (Cov!ST);
!STate6_X with !STate6_Y (Cov!ST);
```

```
!STate7_X with !STate7_Y (Cov!ST);
!STate8_X with !STate8_Y (Cov!ST);
!STate9_X with !STate9_Y (Cov!ST);

!!!!!! Covariance of Global Trait Factor
trait_X with trait_Y (CovGTR); ! Covariance of Global Trait Factor

!!!!!! Covariance & Residual Covariances for auto-regressive (AR) Factors
AR1_X with AR1_Y (CovAR1);  ! T1 covariance , not Residual-Covariance
AR2_X with AR2_Y (RCOVAR);  ! T2 Residual-covariance, not Covariance
AR3_X with AR3_Y (RCOVAR);
AR4_X with AR4_Y (RCOVAR);
AR5_X with AR5_Y (RCOVAR);
AR6_X with AR6_Y (RCOVAR);
AR7_X with AR7_Y (RCOVAR);
AR8_X with AR8_Y (RCOVAR);
AR9_X with AR9_Y (RCOVAR);

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!!!!  Additional Constraints to test Stationarity   !!!!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

model constraint:


!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!!!!!!!Non-linear Inequality Constraints to impose stationarity
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

!!!!! VAR1_X & Y = Variance in auto-regressive (AR) X & Y factor at T1,
!!!!! RVAR_X & Y = Residiual Variance in auto-regressive(AR) X & Y factor at T2-T9;
RVAR_X = VAR1_X-(b_XoX*b_XoX*VAR1_X)-(b_XoY*b_XoY*VAR1_Y)
      -2*(B_XoX*b_XoY*CovAR1);

RVAR_Y = VAR1_Y-(b_YoY*b_YoY*VAR1_Y)-(b_YoX*b_YoX*VAR1_x)
      -2*(B_YoY*b_YoX*CovAR1);

!!!!! CovAR1 = Covariance in auto-regressive (AR) X & Y factor at T1,
!!!!! RCovAR = Residiual CoVariance in auto-regressive (AR) X & Y factor at T2-T9;
RCovAR = CovAR1 *(1 - b_XoX*b_YoY - b_YoX*b_XoY) - (b_XoX*b_YoX*VAR1_X) -
    (b_YoY*b_XoY*VAR1_Y);

!!!!!!! compute three relative variance (RELV) components For X and Y;
!!!!!!  GTR = Global Trait; ST = State; AR = Auto-regressive
new (RelVTR_X,RelVAR_X);  !RelVST_X
 RelVTR_X  = VGTR_X / (VGTR_X+VAR1_X); !VST_X+
! RelVST_X = VST_X/ (VGTR_X+VST_X+VAR1_X);
 RelVAR_X =VAR1_X/ (VGTR_X+VAR1_X); !VST_X+

new (RelVTR_Y,RelVAR_Y); !RelVST_Y
 RelVTR_Y  = VGTR_Y / (VGTR_Y+VAR1_Y);
 !RelVST_Y = VST_Y/ (VGTR_Y+VST_Y+VAR1_Y);
 RelVAR_Y =VAR1_Y/ (VGTR_Y+VAR1_Y);

!!!! New standardized variables representing within-wave covariance terms !!!!!!!
new (GTR_XY, AR_XY);  !, State_XY
GTR_XY = CovGTR/((VGTR_X*VGTR_Y)**.5); !standardized Global Trait (GTR) covariances ;
AR_XY = CovAR1/((VAR1_X*VAR1_Y)**.5); !standardized auto-regressive (AR) covariances ;
!State_XY = CovST/((VST_X*VST_Y)**.5); !standardized STATE (ST) covariances ;

!!!! New variables representing Between-wave paths !!!!!!
```

```
new (stab_X, stab_Y);
  stab_X = B_XoX;
  stab_Y = B_YoY;

new (CPXtoY, CPYtoX);
CPXtoY = b_YoX;
CPYtoX = b_XoY;


!!!!! model estimated longterm relation (LTR) between AR_X at T1 & T9;
new (LTRAR_X,LTRAR_Y);
LTRAR_X = stab_X **9;
LTRAR_Y = stab_Y **9;

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!!!!!!Constrain variance components to be non-negative--often non necessary !!!!!!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

 VGTR_X >0;  VGTR_Y >0;! non-negative global trait X and Yt
 VAR1_X > 0; VAR1_X > 0; !non-negative variance (T1) autor-regressive X and Y;
 RVAR_X > 0; RVAR_X > 0; !non-negative residual variance (T1) autor-regressive X and Y;
! VST_X > 0;   VST_Y > 0; !non-negative STATE variance (X and Y;


Output: sampstat;  stand; SVALUES;
```

Note: By removing the following lines from the Mplus syntax provides a test of partial stationarity (i.e., invariance over time of all autoregressive factors – stability and and cross-lagged paths (constrained to be equal in the longitudinal equilibrium model) and also the within-wave residual variances and covariances (waves 2-9). However, there are no constraints in relation to the variance and covariances in Wave 1 that are specific to the full-trend stationarity test.

```
!!!!! VAR1_X & Y = Variance in auto-regressive (AR) X & Y factor at T1,
!!!!! RVAR_X & Y = Residiual Variance in auto-regressive(AR) X & Y factor at T2-T9;
RVAR_X = VAR1_X-(b_XoX*b_XoX*VAR1_X)-(b_XoY*b_XoY*VAR1_Y)
      -2*(B_XoX*b_XoY*CovAR1);

RVAR_Y = VAR1_Y-(b_YoY*b_YoY*VAR1_Y)-(b_YoX*b_YoX*VAR1_x)
      -2*(B_YoY*b_YoX*CovAR1);

!!!!! CovAR1 = Covariance in auto-regressive (AR) X & Y factor at T1,
!!!!! RCovAR = Residiual CoVariance in auto-regressive (AR) X & Y factor at T2-T9;
RCovAR = CovAR1 *(1 - b_XoX*b_YoY - b_YoX*b_XoY) - (b_XoX*b_YoX*VAR1_X) -
    (b_YoY*b_XoY*VAR1_Y);
```

Section 6
Missing data.

There are overlapping sets of school leaders who completed the survey each year, with some previous participants dropping out and new participants joining the survey. Also, school leaders who missed one year occasionally completed surveys in subsequent years. Hence, our study's total number of leaders was 5,663, but the number who completed surveys each year varied (2019 = 1894; 2018 = 2289; 2017 = 2549; 2016 = 2841; 2015 = 2641; 2014 = 2467; 2013 = 2010; 2012 = 2084; 2011 = 2049).

Here, we applied the full information maximum likelihood method to fully use cases with missing data (FIML; Enders, 2010). FIML results in trustworthy, unbiased estimates for missing values even in the case of large numbers of missing values (Enders, 2010) and is an appropriate method to manage missing data in large longitudinal studies (Jeličič et al., 2009). More specifically, as emphasized in classic discussions of missing data (e.g., Newman, 2014), under the missing-at-random (MAR) assumption that is the basis of FIML, missingness is allowed to be conditional on all variables included in the analyses. However, it does not depend on the values of variables that are missing. This implies that missing values can be conditional on the same variable's values collected in a different wave in a longitudinal panel design. This makes it unlikely that MAR assumptions are seriously violated, as the key situation of not-MAR is when missingness is related to the variable itself. Hence, having multiple waves of parallel data provides strong protection against this violation of the MAR assumption.

The appropriateness of FIML is further strengthened by support for the invariance of parameter estimates over time (see earlier discussion of invariance constraints). Thus, the nature of missing data is like a cohort-sequential design (Marsh, Morin, et al., 2015; Nesselroade & Baltes, 1975; Schaie & Baltes, 1975) in which different cohorts respond at different, overlapping time intervals. Such designs' critical feature is that valid interpretations depend largely on demonstrating that the results are invariant over time (and thus not dependent on the specific cohort). Our cohorts are messy and not as well-demarcated as in a carefully designed cohort-sequential study. However, the clear support for factorial invariance of the measurement model and structural parameters (longitudinal equilibrium, full-stationarity, and partial-stationarity) provides strong evidence that the results generalize over time and, by implication, the multiple missing data patterns.

Nesselroade JR, & Baltes PB. (1979) Longitudinal research in the study of behavior and development. New York: Academic Press.

Newman, D. A. (2014). Missing Data. Organizational Research Methods, 17(4), 372–411. doi.org/1.1177/1094428114548590

Schaie, K.W. & Baltes, B.P. (1975). On sequential strategies in developmental research: Description or Explanation. Human Development, 18: 384-390.

Section 7. The Reliance on Self-report Measures.

We considered self-report measures that might introduce method effects that distort relations between self-efficacy and job satisfaction. This is a difficult issue to address as school leaders are best suited to judge their own self-efficacy and job satisfaction. For RI-CLPMs, method effects that are stable over time are likely to be absorbed into the global (decomposed between-person) trait effects but have little influence on with-person stability and cross-lagged effects. For CLPMs, such method effects are likely to inflate correlations among the constructs, particularly the stability coefficients. For both models, it would be helpful to also collect ratings by colleagues to evaluate self-other agreement and potential method effects using MTMM analyses. Also, we note that our measure of self-efficacy —consistent with much organizational research—focused on general self-efficacy even though it was completed in the context of the work setting (see Marsh, Pekrun, et al., 2019, and discussion in Supplemental Materials Section 1) on the murky distinctions between self-efficacy and other self-belief constructs). Furthermore, it would be useful to expand the outcomes to include objective measures and reports by significant others and evaluate the generalizability of results to non-self-report measures considered here. Nevertheless, we contend that whilst objective measures and reports by observers about leaders self-efficacy and job satisfaction would be a useful addition, these would constitute different constructs to those based on self-reported responses.

Section 8. The Need for Substantive-Methodological Synergy in Contemporary Educational Psychology

To explain our general approach, **l**ike much of our work, this study is substantive-methodological synergy. We try to achieve a constructive balance between the substantive and the methodological foci and explain why the synergy between the two is important. Here we try to make the case for why substantive-methodological synergies are so important for Contemporary Educational Psychology and educational psychological research more generally. Support for this claim can be traced back to the 2007 special issue of Contemporary Educational Psychology devoted to the application of latent variable models in educational psychology and the need for substantive-methodological synergies. In the lead article of this special issue, Marsh and Hau (2007) first coined the term substantive-methodological synergy and presented a manifesto for a "substantive methodological synergy" movement.

In providing an overview for this issue of CEP, Marsh and Hau note that:

> "Distinguishing these articles from those in leading measurement and statistical journals, Kulikowich and Hancock encouraged the submission of manuscripts from authors (or teams) with strong backgrounds in both latent variable methodology and substantive issues in educational psychology. In this respect, they sought to provide an outlet for articles representing a synergy between sophisticated methodology and meaningful substantive issues." (p. 151).

They introduced the topic with the following:

> "As quantitative educational psychologists, we live in exciting times. We are armed with a bevy of new and evolving quantitative tools to address a range of substantive and policy related questions, with statistical power and flexibility that was previously unimaginable. However, this power comes at a cost to the educational psychology research discipline. In order to remain current as researchers—or even informed consumers of this research—we must become conversant with an ever-increasing range of new latent variable statistical procedures. In order to make best use of these new tools, we must pursue research that is at the cutting edge of both latest methodological developments and substantive issues—methodological-substantive synergies. This is not to say, of course, that publications that are

> aimed primarily at methodological concerns or those aimed primarily at substantive issues cannot make important contributions. However, it is our contention that: (a) some of the best methodological research is based on the development of creative methodological solutions to problems that stem from substantive research (e.g., multilevel regression in school/class and student level analyses); (b) new methodologies provide important new approaches to current substantive issues (e.g., latent growth analyses that allow growth data to be collected at varying points on the time line); and (c) methodological-substantive synergies are particularly important in applied areas like educational psychology where single infallible indicators are typically not available." (p. 152)
>
> "Fostering methodological-substantive synergies is difficult. Ideally, these synergistic publications are of interest to both methodologists and substantively oriented researchers. However, with a few notable exceptions, most journals and their readerships are oriented primarily towards one or the other. Furthermore, there may be some potential biases against such synergistic studies even being accepted for publication. In our own experience, for example, we often receive editorial comments that our research is either too substantively oriented (for methodological journals) or too methodologically oriented (for substantive journals)." (p. 152)

After taking the reader through a brief history of selected issues and a review of substantive-methodological synergy articles in that issue of Contemporary Educational Psychology, they concluded:

> "Educational psychology research is experiencing an exciting period, stimulated in part by an explosion of new and evolving latent variable approaches. However, there have not been sufficient demonstrations of the usefulness of these new techniques in substantively meaningful studies. Particularly in an applied area of research like educational psychology, it is important to demonstrate the applicability of new statistical tools in a range of applications that demonstrate their superiority over more traditional approaches. We contend that in order to make the best use of these new tools, we must pursue research that is at the cutting edge of both latest methodological developments and substantive issues—methodological-substantive synergies. However, there are a host of reasons why such synergistic research is not more widespread. From this perspective, the set of studies in this special issue of Contemporary Educational Psychology provides a welcome model, demonstrating how a variety of new latent variable models can be applied to substantively meaningful research issues based on data like that typically encountered in applied educational psychology research. It is our sincere hope that this will encourage other substantive and applied researchers to more fully implement these approaches into their own research programs, and that substantively oriented journals will publish more studies methodological-substantive synergies." (p.168)

The hope of Marsh, Hau, and their colleagues was that the 2007 special issue of Contemporary Educational Psychology would herald a new wave of substantive-methodological synergy that would be embraced by substantively oriented researchers and journals as well as more methodologically oriented researchers and journals. Alas, this appears not to have been the case. Indeed, it seems that the gap between the methodologically oriented and substantively oriented studies continues to be at least as large as in 2007 (e.g., Hoffmans et al., 2021). Nevertheless, we hope the present investigation contributes to bridging this gap in the tradition of Marsh and Hau (2007).

We also emphasize the need for educational psychology research to achieve a constructive balance between methodological and substantive orientations. Support for this claim can be traced back to the 2007 special issue of Contemporary Educational Psychology devoted to the application of latent variable models in educational psychology and the need for substantive-methodological synergies. In the lead article of this special issue, Marsh and Hau (2007, also see SM8) first coined the term substantive-methodological synergy and presented a manifesto for a "substantive methodological synergy" movement. They recognized challenges in this approach, noting substantive-methodological synergies are criticized as either being too substantive (by methodological journals) or too methodological oriented (by substantive journals). From this perspective, the set of studies in the 2007 special issue of Contemporary Educational Psychology provided a welcome model, demonstrating how a variety of new latent variable models can be applied to substantively meaningful research issues based on data like that typically encountered in applied educational psychology research. We hope the present investigation contributes to bridging this gap in the tradition of Marsh and Hau (2007).