Investigating Global Self-Esteem by Integrating Theory and Methods

Dissertation

zur Erlangung des Doktorgrades

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

vorgelegt von

Laura Rebekka Braun

Tübingen

2020

1. Betreuer:	Prof. Dr. Benjamin Nagengast
2. Betreuer:	Dr. Marion Spengler

Tag der mündlichen Prüfung:	22.07.2020
Dekan:	Prof. Dr. Josef Schmid
1. Gutachter:	Prof. Dr. Benjamin Nagengast
2. Gutachter:	Dr. Marion Spengler
3. Gutachter:	Prof. Dr. Steffen Zitzmann

To my cousin Johanna.

Against all odds, your self-esteem was inspiring.

ACKNOWLEDGMENTS

Above all, I would like to thank my supervisory team: Dr. Marion Spengler, Prof. Benjamin Nagengast, Prof. Richard Göllner, Prof. Ulrich Trautwein, and Dr. Sven Rieger. Thank you for giving me this opportunity and for providing me with tremendous academic and emotional support! Thanks for teaching me so much, for believing in me and my ideas, and for encouraging me whenever I needed it. I am grateful that we all got to go through the journey of "Why everybody was wrong (including me)" (Rieger & Nagengast, personal communication) together.

Furthermore, I would like to thank Prof. Brent W. Roberts and Prof. Richard W. Robins for the inspiring and constructive time I was able to spend in their labs in Illinois and California. I gained so much from the time I got to spend with you, and I am beyond grateful to have learned from your expertise. Thanks to Rick for giving me the chance to work with data from the California Families Project.

Very special thanks goes to Cora for her incredible support during every single day of this! Thanks for motivating me over and over again, for being the best feedback buddy out there, and for being a great friend!

Thanks to all of my colleagues and friends at the Hector Research Institute of Education Sciences and Psychology and the LEAD Graduate School and Research Network who supported me throughout this time! In particular, I want to thank the "Personality Group," as well as Johanna, Nele, Tim, Margit, Simone, and Norman. Moreover, I would like to thank Jane and Molly for their language support and for their impressive flexibility.

I am beyond lucky to have amazing friends, who were there when things challenged me the most. In particular, I would like to thank Amelie, Elena, Michi, and Teresa!

Finally, I would not be where I am now without the unconditional support and understanding that my family has given me. Special thanks go to my parents, my sister and my brother, and my "Swabian" family!

ABSTRACT

One major challenge in psychological research is the integration of theory and methods. A successful integration (i.e., the harmonious consideration of theoretical ideas and methodological implementations) is crucial for drawing conclusions from empirical studies back to theoretical phenomena and has the potential to generate synergies for theoretical and methodological progress in science (Greenwald, 2012; Marsh & Hau, 2007). However, long-cherished assumptions in well-established research fields, accompanied by growing analytical complexity, have often limited the potential for substantive-methodological integrations. Therefore, the present dissertation was aimed at integrating theory and methods for one of the most well-studied constructs in psychology, namely, the global self, oftentimes represented as the construct *global self-esteem*. Global self-esteem describes individuals' overall subjective feelings of worth and has attracted the interest of many researchers due to its relevance in the context of mental health (e.g., James, 1890/1963; Orth & Robins, 2014; Rosenberg, 1989). Studying self-esteem is of particular concern during adolescence and young adulthood because, in this phase, individuals have to face many developmental and environmental challenges (Harter, 1998; Rosenberg, 1986). Global self-esteem has been described as a construct that is *unidimensional* (e.g., Rosenberg, 1989), *trait-like* (e.g., Orth & Robins, 2019), and socially manifested through *parents and peers* (e.g., Cooley, 1902; Harris & Orth, 2019; Leary & Baumeister, 2000). At the same time, however, there are deviations from and extensions of these assumptions such as conceptualizations of *multidimensional, hierarchical global self-concept* (Shavelson, Hubner, & Stanton, 1976), a consideration of *state-like* self-esteem (Leary & Baumeister, 2000), and the incorporation of *other social contexts* beyond parents and peers (e.g., students' interactions with teachers in classrooms). Despite the theoretical relevance of these deviations and extensions, they have received only a little empirical attention in research on global self-esteem. One reason for this gap could be that sophisticated methodological implementations for these research foci have been missing.

The present dissertation was thus aimed at integrating theory and methods in research on self-esteem. Thereby, this dissertation pursued two overarching objectives. The first objective was to improve the understanding of self-esteem in adolescence and young adulthood. For this purpose, this dissertation (a) addressed different conceptualizations of global self-concept as the apex of a multidimensional hierarchy and brought them together with global self-esteem, (b) investigated state and trait self-esteem and the consequences for their relations with depressive symptoms, and (c) examined reciprocal relations between self-esteem and student-teacher

relationships and examined these relations over time. The second objective was to improve the understanding of different methodological implementations (namely, the operationalization of higher order constructs, states and traits, and reciprocal relations) and their unique consequences for the aforementioned research questions and beyond. The two overarching objectives of this dissertation were addressed in three empirical studies.

In the first study (*Rethinking the Elusive Construct of Global Self-Concept: A Latent Composite Score as the Apex of the Shavelson Model*), different conceptualizations of the global self (i.e., global self-concept and global self-esteem) were examined. As the focus of the study, two different conceptualizations of global self-concept as the apex of the multidimensional self-concept hierarchy were compared by applying different analytical models to represent higher order constructs. Using three independent large-scale studies (N_1 = 8,068; N_2 = 3,876; N_3 = 2,095) of adolescents and young adults, we (a) applied second-order factor models, which assume that global self-concept affects lower order domain-specific self-concepts, and (b) compared them with a model-based latent composite scores, which reflect processes by which global self-concept is formed on the basis of domain-specific self-concepts. Our results indicated advantages of the latent composite scores as indicated by higher variances and a more plausible pattern of stabilities and correlations with external criteria, such as unidimensional global self-esteem. Therefore, we propose that global self-concept—the apex of the multidimensional hierarchy of self-concept—be modeled as a latent composite score. Over and above this, the study indicated that the conceptualization of *multidimensional hierarchical global self-concept* was more aligned with *unidimensional global self-esteem* when nonacademic self-concepts were included in comparison with academic self-concepts.

In the second study (*How State and Trait Versions of Self-Esteem and Depressive Symptoms Affect Their Interplay: A Longitudinal Experimental Investigation*), the stability of self-esteem was addressed by using a two-fold operationalization of states and traits (measurement and modeling approach). Using exploratory (N_1 = 683) and preregistered confirmatory (N_2 = 1,087) experimental longitudinal designs, university students were randomly assigned to state and trait measures of self-esteem (and depressive symptoms), which were operationalized by using different time frames in the questionnaires (“In general…” vs. “During the last 2 weeks…”). The results indicated that, first, consistently across the two studies, the trait time frames revealed higher proportions of trait variance and lower proportions of state residual variances than the state time frames. Second, across the two studies, the cross-lagged relations between self-esteem and depressive symptoms depended on the time frames used to assess the

constructs. Third, when controlling for stable trait differences, the cross-lagged results were least consistent when trait time frames were used, which highlighted the interdependency involved in measuring and modeling state and trait self-esteem.

In the third study (*Is Teacher Attachment Prospectively Related to Self-Esteem? A 10-Year Longitudinal Study of Mexican-Origin Youth*), the reciprocal relation between student-teacher relationships and students' self-esteem was investigated by using two different cross-lagged panel models. The study used data from $N = 674$ Mexican-origin students followed annually from age 11 to 21 and tested eight preregistered hypotheses about reciprocal relations between students' perceived teacher attachment (i.e., teacher support and teacher rejection) and students' global self-esteem. The results indicated (a) positive prospective reciprocal relations between teacher support and self-esteem, based on cross-lagged panel models (CLPMs; focus on overall between-person differences) as well as random-intercept cross-lagged panel models (RI-CLPMs; focus on differences in within-person deviations), and (b) negative prospective reciprocal relations between teacher rejection and self-esteem, based only on CLPMs but not on RI-CLPMs. Overall, the results suggested that transactional processes underlie reciprocal relations between teacher attachment and self-esteem, whereas the results were more consistent in the CLPMs than in the RI-CLPMs.

From a theoretical perspective, this dissertation refines the understanding of (a) the relation between unidimensional global self-esteem and multidimensional, hierarchical global self-concept, (b) trait *and* state self-esteem, as well as (c) individual and environmental predictors and consequences of self-esteem. From a methodological perspective, across the three studies, this dissertation observed important empirical differences from *different* methodological implementations. Thereby, this dissertation points to the consequences of cross-sectional and longitudinal higher order factor models and emphasizes the importance of integrating theory, methods, *and* data.

ZUSAMMENFASSUNG

Eine der größten Herausforderungen in der psychologischen Forschung ist die Integration von Theorie und Methoden. Eine erfolgreiche Integration (d.h. die aufeinander abgestimmte Berücksichtigung theoretischer Ideen und methodischer Umsetzungen) ist entscheidend, um Schlussfolgerungen aus empirischen Studien auf theoretische Phänomene zurückführen zu können, und hat das Potenzial, synergetische Erkenntnisse für den theoretischen und methodischen wissenschaftlichen Fortschritt zu generieren (Greenwald, 2012; Marsh & Hau, 2007). Lang gehegte Annahmen in etablierten Forschungsfeldern und wachsende Komplexität methodischer Analyseverfahren schränken jedoch häufig das Potenzial für theoretisch-methodische Integrationen ein. Das Anliegen der vorliegenden Dissertation war es daher, Theorie und Methoden für eines der am besten erforschten Konstrukte in der Psychologie zu integrieren, nämlich für das globale Selbst, das oft als *globales Selbstwertgefühl* bezeichnet wird. Das globale Selbstwertgefühl beschreibt den subjektiv wahrgenommen Wert der eigenen Person und hat aufgrund seiner Relevanz im Kontext psychischer Gesundheit das Interesse vieler Forscherinnen und Forscher geweckt (z.B. James, 1890/1963; Orth & Robins, 2014; Rosenberg, 1989). Die Erforschung des globalen Selbstwertgefühls scheint besonders wichtig während der Phase der Adoleszenz und des jungen Erwachsenenalters, da Individuen in dieser Zeit mit besonders vielen entwicklungsbedingten und kontextuellen Herausforderungen konfrontiert sind. Das globale Selbstwertgefühl wird als ein Konstrukt beschrieben, das *eindimensional* (z.B. Rosenberg, 1989), *trait-like* (d.h. eine eher stabile Eigenschaft; z.B. Orth & Robins, 2019) und *durch Eltern und Peers sozial manifestiert* ist (z.B. Cooley, 1902; Harris & Orth, 2019; Leary & Baumeister, 2000). Gleichzeitig bestehen aber auch Abweichungen und Erweiterungen dieser Annahmen, wie z.B. die Konzeptualisierungen eines *multidimensionalen, hierarchischen globalen Selbstkonzepts* (Shavelson, Hubner, & Stanton, 1976), die Berücksichtigung eines *state-like* Selbstwertgefühls (d.h. eher situative Zustände; Leary & Baumeister, 2000) und die Einbeziehung *anderer sozialer Kontexte* jenseits von Eltern und Peers (z.B. Interaktionen von Schülerinnen und Schülern mit ihren Lehrkräften im Klassenzimmer). Trotz der theoretischen Relevanz dieser Abweichungen und Erweiterungen haben sie in der Forschung zum globalen Selbstwertgefühl nur wenig empirische Aufmerksamkeit erhalten. Ein Grund für diese Forschungslücken könnte sein, dass angemessene methodische Implementierungen dieser Forschungsschwerpunkte fehlten.

Die vorliegende Dissertation hatte daher zum Ziel, Theorie und Methoden in der Forschung zum globalen Selbstwertgefühl zu integrieren. Dabei verfolgte diese Dissertation zwei

übergeordnete Ziele. Das erste Ziel bestand darin, das theoretische Verständnis über das Selbstwertgefühl im Jugend- und jungen Erwachsenenalter zu verbessern. Aus diesem Grund untersuchte diese Dissertation (a) verschiedene Konzeptualisierungen des globalen Selbstkonzepts als Spitze einer multidimensionalen Hierarchie sowie deren Zusammenhang mit dem eindimensionalen globalen Selbstwertgefühl, (b) das Trait- und State-Selbstwertgefühl und deren Konsequenzen für den Zusammenhang mit depressiven Symptomen, und (c) den längsschnittlichen Zusammenhang zwischen dem globalen Selbstwertgefühl und Schüler-Lehrkraft-Beziehungen im Verlauf der Schulzeit und darüber hinaus. Das zweite Ziel dieser Dissertation war die Verbesserung des Verständnisses verschiedener methodischer Implementierungen (und zwar: die Operationalisierung von Konstrukten höherer Ordnung, von States und Traits sowie Modelle zur Analyse reziproker Zusammenhänge) und ihre spezifischen Konsequenzen für die oben genannten Forschungsfragen sowie über das Selbstwertgefühl hinaus auch für weitere Forschung. Die beiden übergeordneten Ziele dieser Dissertation wurden in drei empirischen Studien adressiert.

In der ersten Studie (*Rethinking the Elusive Construct of Global Self-Concept: A Latent Composite Score as the Apex of the Shavelson Model*) wurden verschiedene Konzeptualisierungen des globalen Selbst (d.h. globales Selbstkonzept und globales Selbstwertgefühl) untersucht. Im Mittelpunkt der Studie standen zwei verschiedene Konzeptualisierungen des globalen Selbstkonzepts als Spitze einer multidimensionalen Selbstkonzepthierarchie, die durch die Anwendung verschiedener analytischer Modelle zur Darstellung von Konstrukten höherer Ordnung verglichen wurden. Unter Verwendung von drei unabhängigen Large-Scale Studien basierend auf Daten von Jugendlichen und jungen Erwachsenen (N_1 = 8,068; N_2 = 3,876; N_3 = 2,095) wurden zwei Ansätze miteinander vergleichen: (a) Faktormodelle zweiter Ordnung, die davon ausgehen, dass das globale Selbstkonzept domänenspezifische Selbstkonzepte niedrigerer Ordnung beeinflusst und (b) modellbasierte latente Composite Scores, denen die Annahme zugrunde liegt, dass sich das globale Selbstkonzept auf Grundlage domänenspezifischer Selbstkonzepte formiert. Die Ergebnisse wiesen auf Vorteile der latenten Composite Scores hin, welches sich in höheren Varianzen und einem plausibleren Muster an Stabilitäten und Korrelationen mit externen Kriterien, wie z.B. dem Zusammenhang mit dem eindimensionalen globalen Selbstwertgefühl, zeigte. Daher wird vorgeschlagen, das globale Selbstkonzept—die Spitze der multidimensionalen Selbstkonzepthierarchie—auch in zukünftiger Forschung als latenten Composite Score zu modellieren. Die Studie wies außerdem darauf hin, dass das globale

Selbstkonzept (basierend auf der multidimensionalen Hierarchie) stärker mit dem eindimensionalen globalen Selbstwertgefühl zusammenhängt, wenn nicht-akademische Selbstkonzepte im Vergleich zu akademischen Selbstkonzepten berücksichtigt wurden.

In der zweiten Studie (*How State and Trait Versions of Self-Esteem and Depressive Symptoms Affect Their Interplay: A Longitudinal Experimental Investigation*) wurde die Stabilität des Selbstwertgefühls anhand einer zweiteiligen Operationalisierung (Mess- und Modellierungsansatz) von States und Traits untersucht. Unter Verwendung einer explorativen (N_1 = 683) und einer präregistrierten konfirmatorischen (N_2 = 1.087) experimentellen, längsschnittlichen Studie wurden Studierende randomisiert State- und Trait-Messungen des Selbstwertgefühls (und depressiver Symptome) zugewiesen, welche durch die Verwendung unterschiedlicher Zeitreferenzen in den Fragebögen operationalisiert wurden ("Im Allgemeinen..." vs. "Während der letzten 2 Wochen..."). Die Ergebnisse deuteten darauf hin, dass die Trait-Messungen in beiden Studien konsistent höhere Anteile an Trait-Varianz (zeitstabile Varianz) und niedrigere Anteile an State-Residualvarianz (zeitpunktspezifische Varianz) aufwiesen als die State-Messungen. Des Weiteren waren die längsschnittlichen Zusammenhänge zwischen dem Selbstwertgefühl und depressiven Symptomen über beide Studien hinweg von der zeitlichen Referenz (State vs. Trait) der Messungen abhängig. Die Ergebnisse aus Cross-Lagged-Panel-Modellen, die für zeitstabile Unterschiede kontrollieren, waren am wenigsten konsistent wenn Trait-Messungen verwendet wurden, was die Interdependenz bei der Messung und Modellierung von Trait- und State-Aspekten des Selbstwertgefühls deutlich machte.

In der dritten Studie (*Is Teacher Attachment Prospectively Related to Self-Esteem? A 10-Year Longitudinal Study of Mexican-Origin Youth*) wurde der reziproke Zusammenhang zwischen dem Selbstwertgefühl von Schülerinnen und Schüler und der Schüler-Lehrkraft-Beziehung (d.h. der Bindung von Schülerinnen und Schülern zu ihren Lehrkräften) mit Hilfe zweier unterschiedlicher Cross-Lagged-Panel-Modelle untersucht. Die Studie verwendete Daten von N = 674 in den USA lebenden Jugendliche mexikanischer Herkunft, die jährlich ab dem Alter von 11 bis 21 Jahren befragt wurden, und testete acht präregistrierte Hypothesen über den Zusammenhang zwischen dem Selbstwertgefühl der Schülerinnen und Schüler und der Schüler-Lehrkraft Beziehung (erfasst über Schülerratings). Die Ergebnisse zeigten (a) positive reziproke Zusammenhänge zwischen Selbstwertgefühl und wahrgenommener Lehrkraft-*unterstützung* sowohl basierend auf klassischen Cross-Lagged-Panel-Modellen (CLPMs; Betrachtung von allgemeinen Unterschieden zwischen Personen) als auch basierend auf Random Intercept Cross-Lagged-Panel-Modellen (RI-CLPMs; Betrachtung von Unterschieden in den

Abweichungen innerhalb von Personen), und (b) negative reziproke Zusammenhänge zwischen Selbstwertgefühl und wahrgenommener Lehrkraft*ablehnung* basierend auf CLPMs, nicht aber auf RI-CLPMs. Insgesamt legten die Ergebnisse nahe, dass transaktionale Prozesse den reziproken Beziehungen zwischen Selbstwertgefühl und Schüler-Lehrkraft-Beziehung zugrunde liegen, während die Ergebnisse in den CLPMs konsistenter waren als in den RI-CLPMs.

Aus theoretischer Perspektive verbessert diese Dissertation das Verständnis (a) des Zusammenhangs zwischen eindimensionalem, globalen Selbstwertgefühl und multidimensionalem, hierarchischen globalen Selbstkonzept, (b) des Trait (eher eigenschaftsähnlichem) *und* State (eher situativen) Selbstwertgefühls sowie (c) von individuellen und kontextuellen Prädiktoren und Konsequenzen des globalen Selbstwertgefühls. Aus methodischer Sicht wurden in allen drei Studien dieser Dissertation wichtige empirische Unterschiede *verschiedener* methodischer Umsetzungen beobachtet. Dabei weist diese Dissertation auf die Konsequenzen von quer- und längsschnittlichen Faktorenmodellen höherer Ordnung hin und betont die Bedeutung der Integration von Theorie, Methoden *und* Daten.

CONTENTS

1 INTRODUCTION AND THEORETICAL FRAMEWORK.....1

1.1 The Global Self 4

1.1.1 Theoretical Origins of the Self 4

1.1.2 Different Conceptualizations of the Global Self 8

1.2 The Development of Global Self-Esteem 11

1.3 Predictors and Consequences of Global Self-Esteem 14

1.3.1 Individual Predictors and Consequences.....15

1.3.2 Environmental Predictors and Consequences 18

1.4 Recapping the Questions that have Emerged about Self-Esteem 22

1.5 Integrating Theoretical and Methodological Considerations.....24

1.5.1 Higher Order Constructs 26

1.5.2 States and Traits 30

1.5.3 Reciprocal Relations 34

2 AIMS AND RESEARCH QUESTIONS.....43

3 STUDY 1: Rethinking the Elusive Construct of Global Self-Concept: A Latent Composite Score as the Apex of the Shavelson Model**.....47**

4 STUDY 2: How State and Trait Versions of Self-Esteem and Depressive Symptoms Affect Their Interplay: A Longitudinal Experimental Examination **.....95**

5 STUDY 3: Is Teacher Attachment Prospectively Related to Self-Esteem? A 10-Year Longitudinal Study of Mexican-Origin Youth **.....177**

6 GENERAL DISCUSSION 225

6.1 Theoretical Advances for Research on Self-Esteem 227

6.1.1 The Relation between Global Self-Concept and Global Self-Esteem 227

6.1.2 State and Trait Self-Esteem 229

6.1.3 Relations between Self-Esteem and Individual and Environmental Factors 231

6.2 Advances from and for Integrating Methodological Perspectives 233

6.2.1 A Common Thread: Cross-Sectional and Longitudinal Higher Order Factor Models 233

6.2.2 Integrating Theory, Methods, *and* Data 235

6.3 Limitations and Future Directions 237

6.3.1 Theoretical Limitations and Future Directions 237

6.3.2 Methodological Limitations and Future Directions 239

6.4 An Outlook: A Time Frame – Specificity Classification 241

6.5 Implications for the Construct of Global Self-Esteem in Policy and Practice 245

6.6 Conclusion 247

7 REFERENCES 249

1 INTRODUCTION AND THEORETICAL FRAMEWORK

How do we generate scientific knowledge about the constructs that are central to our research? This might be one of the broadest questions for which every area of science has found its unique nuanced answer. In psychological research, this question is an ever-present challenge, given that many psychological constructs are unobservable, hypothetical constructs (Cronbach & Meehl, 1955). Therefore, psychological researchers, in particular, need to properly define their theoretical constructs and subsequently draw on the *most appropriate* methodological operationalization of the theoretical phenomena of interest (Borsboom, Mellenbergh, & van Heerden, 2003; Borsboom, 2006; Marsh & Hau, 2007). In this regard, not only is it self-evident that theories and methods are connected, but their link is imperative because a methodological choice needs to be made on the basis of the theoretical definition of a research question. However, long-cherished assumptions in well-established research fields and growing analytical challenges have oftentimes produced a standstill in in-depth considerations about the link between theory and methods (Marsh & Hau, 2007). By contrast, research integrating theoretical and methodological considerations has the potential to generate synergies for theoretical and methodological scientific progress (Greenwald, 2012; Marsh & Hau, 2007). The present dissertation attempts to approach such "substantive-methodological synergies" (Marsh & Hau, 2007, p. 151) in a well-established psychological research field, namely, the field of global self-esteem. This joint venture needs to begin by delving into the theoretical foundations and theoretical stances of the respective gaps in research, followed by considering the most appropriate methodological representations.

Global self-esteem is one of the oldest constructs in psychological research and has been studied intensively across a broad range of psychological disciplines such as personality, educational, developmental, social, and clinical psychology (e.g., Donnellan, Trzesniewski, & Robins, 2011; Trzesniewski, Donnellan, & Robins, 2013). Global self-esteem is defined as the "individual's subjective evaluation of his or her worth as a person" (Donnellan et al., 2011, p.718). Among a broader set of socioemotional skills (e.g., Big Five personality traits, life satisfaction, motivation), self-esteem has captured the interest of many researchers, policymakers, and therapists, primarily due to its relations to psychological indicators of mental health (Sowislo & Orth, 2013; Trzesniewski et al., 2006). The emergence and development of self-esteem is of particular interest during adolescence and young adulthood because in this phase, individuals face a broad range of developmental and environmental changes (Harter, 1998; Rosenberg, 1986). On the basis of a comprehensive research field, the majority of researchers

have suggested that self-esteem reflects a construct that is (a) *unidimensional* (e.g., Rosenberg, 1989), (b) *trait-like* (e.g., Orth & Robins, 2019), and (c) socially shaped by *parents and peers* (e.g., Cooley, 1902; Harris & Orth, 2019; Leary & Baumeister, 2000). Despite researchers largely building upon these three global assumptions about self-esteem, deviations and extensions from these long-cherished assumptions have not been addressed sufficiently. In fact, urgent subsequent questions involve how previous conceptions of global self-esteem are related to (a) *multidimensional, hierarchical* global self-perceptions (Shavelson et al., 1976), (b) *state-like* conceptions of self-esteem (Leary & Baumeister, 2000), and (c) *other social contexts* beyond parents and peers (e.g., students' interactions with teachers in classrooms). Besides a theoretical paucity, these urgent questions face important methodological challenges, such as how to model a higher order construct, how to operationalize states and traits, and how to choose analytical models to capture reciprocal relations. Different methodological implementations of these challenges are fundamentally related to theoretical assumptions regarding self-esteem. Therefore, studying these crucial questions about self-esteem requires a closer link between theoretical and methodological considerations.

Accordingly, I aim to address two overarching objectives with this dissertation. The first objective is to improve the theoretical understanding of self-esteem. More specifically, this dissertation investigates the conceptualization of the multidimensional hierarchical global self-concept, the stability of state and trait self-esteem, as well as the relations between student-teacher relationships and self-esteem. The second objective is to improve the understanding of specific methodological implementations and their consequences over and above self-esteem by revisiting *different* methodological approaches for hierarchical constructs, states and traits, as well as reciprocal relations over time and set them in relation to broader theoretical assumptions as well as empirical consequences.

This dissertation is structured as follows: The theoretical background is divided into two broad parts. In the first broad part, I introduce the theoretical foundations of self-esteem by reviewing the theoretical origins of self-esteem (Chapter 1.1), empirical findings on the development of self-esteem (Chapter 1.2), and individual and environmental predictors and consequences (Chapter 1.3). Subsequently, I summarize this first part and identify three emerging areas of interest (Chapter 1.4). In the second broad part, I merge these areas of interest with their respective methodological challenges (Chapter 1.5). In doing so, I scrutinize the operationalizations of hierarchical constructs, states, and traits, and reciprocal relations. From both the substantive and methodological perspectives, I derive the research questions presented in

the three empirical studies included in this dissertation (Chapter 2). Subsequently, I present the three empirical studies, which represent the main part of this dissertation (Chapters 3 to 5). Finally (Chapter 6), the findings from the empirical studies are discussed with regard to their relevance for theory and methods and with respect to limitations, future research, and implications for practice and policy. The dissertation closes with an overall conclusion.

1.1 The Global Self

Research on the self has a long history, which fundamentally began with the ideas presented by William James (1890/1963) and split into different theoretical streams throughout the last century. In order to provide a comprehensive picture of its theoretical grounding, I want to provide an overview of some of the most influential theoretical perspectives. This begins with *James' perspective*, followed by multiple theoretical streams that I will sum up under the term *social perspectives*. Subsequently, I will present an *attitude perspective* and approaches that embrace *a multidimensional perspective*. It is important to note that these perspectives are not necessarily opposed to each other, but they all place a different focus on the emergence and nature of the global self, which is discussed further at the end of this chapter.

1.1.1 Theoretical Origins of the Self

Research on the *self* goes back to the American psychologist William James, who asked what constitutes our views of our selves. James (1890/1963) divided the self into two aspects: The self as the knower ("I"), which actively takes control of one's own perceptions and behaviors; and the self as a known ("Me"), which reflects the person's self-views. Thus, the "Me" component refers to what we consider perceptions of the self (e.g., self-esteem and self-concept). James emphasized the complex environment in which individuals act ("I") and perceive ("Me"). This is why he proposed three hierarchical levels of the self: (a) the material self, which comprises a person's appearance, clothes, house, and other possessions, (b) the social self, which represents the extent to which a person is appreciated by others in the environment, and (c) the spiritual self, which reflects inner states such as thoughts and dispositions. According to James, these levels of the self are hierarchically ordered with the material self at the bottom, the social self on the intermediate level, and the spiritual self at the highest level. James further emphasized that different components were not all relevant to the same degree within and across individuals. By contrast, the relevance of the respective components for the individuals' perceptions of the self results from a ratio of success and aspirations in different areas of life. As such, James argued that when success exceeds a person's own aspirations, this will increase the person's overall feeling of the self.

Other theoretical perspectives, which initially evolved around the same time as James' formulations, have pointed to the superior relevance of the social environment in the construction of the self-view. These perspectives will be subsumed under the *social perspective*. Among the most prominent approaches is symbolic interactionism, which goes back to Cooley (1902).

Cooley described the “looking glass self” as a metaphor, which referred to others’ judgments as a social mirror that serves as the main source for the self-view. Thus, for Cooley, individuals internalize what they believe others think of them, such as about their appearance, characteristics, and attitudes. Cooley noted that this dependence on social sources decreases with age. Even if initially shaped by the social environment, adults most often develop mature and stable self-views that are less contingent upon momentary shifts in their social mirror. Mead (1934) and Baldwin (1895) shared Cooley’s ideas about the construction of the self in the context of social interactions. According to Mead (1934), individuals adopt a “generalized other” perspective of the self, which is less determined by specific others and more determined by the general view of the social environment. Baldwin (1895) emphasized the role of social imitation processes in infancy that contribute to individuals’ self-views. Thus, children internalize the behavior of their caregivers and incorporate it into their sense of self. Similarly, attachment theorists (e.g., Ainsworth, Blehar, Waters, & Wall, 1978; Bowlby, 1969, 1973, 1980) proposed that early attachment to caregivers shapes people’s internal working models and that these internal working models form the foundation of individuals’ self-perceptions. Largely in line with assumptions from symbolic interactionists and attachment theorists, sociometer theory by Leary and Baumeister (2000) builds upon the social relevance of the self. According to sociometer theory, *self-esteem* reflects a person’s subjectively perceived relational value, which is “the degree to which other people regard their relationships with the individual as valuable or important” (Leary, 2004, p. 375). Hence, self-esteem functions as a sociometer that assesses the quality of social relationships and reinforces behavior that helps maintain or increase the sociometer. Metaphorically, Leary and Baumeister (2000) compared self-esteem to a fuel gauge, which monitors the fuel level and alerts the individual when there is a lack of fuel. In the wake of this theory, self-esteem changes when the individual perceives a shift in his or her relational value. These shifts might occur on the one hand through objective changes in one’s standing on social attributes (e.g., friendliness, trustworthiness, or social status). On the other hand, a person’s relational value can change due to changes in processing the information on social attributes (e.g., selective attention, the selection of specific social contexts, or the reweighting of certain social attributes). Leary and Baumeister (2000) further distinguished two qualitatively different monitor systems where these shifts take place. *Trait self-esteem* monitors the relatively enduring relational value, which reflects a cross-situational perception of acceptance and rejection and is largely resistant to social feedback from specific situations. By contrast, *state self-esteem* refers to a current relational value, which represents short-term feelings of worth and can vary across situations.

Along with agreeing that the social situation is relevant for self-esteem, the sociologist Morris Rosenberg (1965, 1979, 1989) developed an *attitude perspective* on self-esteem by comparing the attitude toward the self with an attitude toward an object. More specifically, he suggested that all relevant dimensions of attitudes—including content, direction, intensity, importance, salience, consistency, clarity, and stability—are dimensions of a person's attitude toward the self (Rosenberg, 1989). Despite large similarities between attitudes toward an object and toward the self, Rosenberg also noted that self-attitudes have some unique qualities in comparison with other attitudes, such as that everyone is motivated to exhibit the same positive attitude or that more emotional reactions are involved. Nevertheless, given the large degree of comparability, Rosenberg proposed that measures that were used to assess attitudes toward objects be transferred to the measurement of attitude toward the self. Accordingly, he constructed the Rosenberg Self-Esteem Scale (RSE; Rosenberg, 1965, 1989), which is a generalized, unidimensional measure of global self-esteem. The measure comprises 10 items that are designed to assess a person's attitude toward the self with items such as "All in all, I am satisfied with myself." To this day, this has been the state-of-the-art and most widely used instrument for assessing global self-esteem (Donnellan, Trzesniewski, & Robins, 2015). Rosenberg emphasized that the RSE is able to measure whether individuals perceive themselves as people of worth, which, by contrast, does not imply or measure whether individuals feel superior to others. Rosenberg (1989) also pointed out the social influences of self-esteem, which he emphasized not only on the level of significant others (e.g., parents) but also on a more global societal level (i.e., through a person's social class or in a religious context). Rosenberg further noted that the phase of adolescence is particularly important for the development of self-esteem (1986) because this phase is marked by important changes in the ability to perceive how others see you and in the ability to cognitively evaluate and describe yourself in different areas of life. Rosenberg (1989) emphasized that low self-esteem is related to feelings of anxiety and depressive symptoms and shapes how individuals behave in their social lives and in society. Furthermore, Rosenberg (1986) distinguished between barometric self-esteem, which describes short-term fluctuations, and baseline self-esteem, which refers to enduring, slowly changing levels.

In contrast to the social and attitude perspectives, which usually focus on the unidimensional, overall perception of the self and usually embrace the term *self-esteem*, other theoretical approaches have more directly expanded on James' initial ideas about a multidimensional and hierarchically ordered perspective on the self, which is what I refer to as the *multidimensional perspectives*. These perspectives usually use the term *self-concept* when referring to domain-

specific aspects of the self but also when referring to global (or general) components of the self (e.g., Harter, 1999; Marsh, 1987b; Shavelson et al., 1976).[1] One of these approaches is the developmental framework created by Susan Harter (e.g., 1983, 1990, 1998) who focused on multidimensional, domain-specific self-concepts and the construction of global self-concept. Harter built her theory on the developmental stages proposed by Piaget (1960) and described the characteristics of self-concepts in six developmental stages ranging from early childhood to late adolescence. Harter (1998) emphasized that children's and adolescents' self-views change with increasing age due to increases in their abilities in the differentiation, abstraction, introspection, and integration of different sources of knowledge. Harter proposed a bottom-up process for the development of global aspects of the self, in which domain-specific self-concepts, particularly the self-concepts that are considered to be important for the individual, predict global self-esteem (see also James, 1890/1963). Moreover, Harter corroborated Cooley's assumptions (1902) about the significance of the social environment in shaping self-perceptions (Harter, 1998). More specifically, Harter underlined the relevance of parents early in life and the growing importance of classmates and teachers during the school years for the development of self-esteem in youth and adolescence. Furthermore, Harter closely linked low self-esteem to affect and general hopelessness, which together have been found to constitute a composite indicator of depression (Harter, 1993, 1998; Harter, 1999). The most influential multidimensional model of self-concept was posed by Shavelson et al. (1976), who described a multidimensional, hierarchical self-concept (Marsh & Shavelson, 1985; Shavelson et al., 1976). Similar to Harter's approach, Shavelson et al.'s approach takes into account both domain-specific self-concepts and a global self-concept. Shavelson et al. proposed that self-concept is multidimensional (e.g., math self-concept, social self-concept) and hierarchically ordered with domain-specific self-concepts at lower levels of the hierarchy, and global self-concept at the apex of this hierarchy. Shavelson et al. also proposed higher stability for components that are more global (i.e., higher in the hierarchy), whereas the more specific a self-concept is (i.e., the lower it is located in the hierarchy), the less stable it should be. At the same time, they proposed that self-concept is developmental and that domain-specific self-concepts become more differentiated across childhood and adolescence.

[1] Initially, a distinction between self-concepts and self-esteem was made in reference to the descriptive nature of self-concepts and the evaluative nature of self-esteem. However, this distinction has largely been revoked due to the assumption that both domain-specific self-concepts and global self-esteem are descriptive *and* evaluative in nature (e.g., Harter, 1999; Shavelson et al., 1976).

In sum, the theoretical origins of self-esteem have a long history, thus laying the foundation for research on global as well as domain-specific evaluations of the self.[2] The basic assumptions by James as well as the social, attitude, and multidimensional perspectives agree that the global self develops across the lifespan and that there is particular potential for malleability during adolescence. The perspectives differ slightly in their assumptions about the degree of stability of the global self: Whereas the Shavelson model describes global self-concept as relatively stable, Leary and Baumeister (2000) and Rosenberg (1986) emphasized the difference between enduring trait self-esteem and more malleable state self-esteem. Across all theoretical streams, the social context plays an important role in shaping individuals' self-esteem, as self-perceptions are supposed to manifest in interactions with others. Obviously, the social perspectives have placed greater emphasis on social relationships than the other approaches. According to the social perspectives, self-esteem is a social mirror of attachment or of positive interactions with significant others. At the same time, previous approaches have differed in whether they have focused primarily on a unidimensional global self-esteem or whether they have emphasized multidimensional and hierarchical levels of self-concepts. Therefore, there have been divergent understandings of the global self, ranging from a global self-esteem as an attitude that can be measured directly (Rosenberg, 1979) to global self-concept at the apex of the self-concept hierarchy (Shavelson et al., 1976). In the following, I will narrow the view to the different definitions of the global self on the basis of the different theoretical origins.

1.1.2 Different Conceptualizations of the Global Self

Previous research has used diverging conceptualizations of the global self, most importantly differing with respect to assumptions about the dimensionality and measurement thereof. According to Marsh and Shavelson (1985), three different conceptualizations of the global self have circulated and can still be identified in current research: The first and most prominent definition is the directly measured construct of *global self-esteem*, most prominently assessed by the Rosenberg Self-Esteem Scale (RSE; Rosenberg, 1979, 1989). According to this approach, the global self can be measured by asking participants to rate their global feelings of the self. Along with the RSE, there are a variety of other scales that measure the global self directly, and these measures typically demonstrate high correlations with the RSE (e.g., Marsh, 1992; Robins, Hendin, & Trzesniewski, 2001; Tafarodi & Swann, 2001; for an overview of

[2] It is important to note that there are other theoretical perspectives on self-esteem, such as more cognitive-oriented approaches (e.g., Epstein, 1973; Markus, 1977), which are not the focus of the present dissertation.

different measures, see Donnellan et al., 2015). This approach has been a particularly strong force in guiding research on social and personality psychology (Donnellan et al., 2015) and has typically applied the term *global self-esteem* to describe the construct. Rather than measuring the global self directly, the other two conceptualizations were based on multidimensional perspectives on the self and created an indicator of the global self by using domain-specific self-concept items. Here, they used the term *global (or general) self-concept* instead of global self-esteem. Hence, in the second definition, global self-concept was operationalized as a total score (e.g., a sum) composed of a variety of domain-specific self-concept items. Before the Shavelson model was introduced, this was a very common way to model global self-concept (e.g., Coopersmith, 1967; Fitts, 1965). But, at that time, the selection of items as well as the analytical approaches lacked theoretical and methodological grounding. The third definition of global self-concept goes back to Shavelson et al. (1976), who proposed a more structured theoretical framework for self-concept that was based on the multidimensional, hierarchical model described above, in which global self-concept represented the apex of the hierarchy. Theoretically, they assumed a *bottom-up* process through which global self-concept was formed by appraising multiple characteristics, a procedure that is in line with assumptions put forth by James (1890/1963) and Harter (1990, 1998). Empirically, they applied second-order factor models (Marsh & Hocevar, 1985; e.g., Marsh & Shavelson, 1985; Marsh, 1987b, 1990), which are more aligned with a *top-down* process. However, the second-order factors did not fit the data well, most likely because the different domain-specific self-concepts had only low intercorrelations. This pattern of results was later described in the framework of dimensional comparison processes as indicating that individuals contrast their own self-concepts across different domains (e.g., Marsh, 1986; Marsh & Hau, 2004). Based on these findings, further developments on the hierarchy of self-concept tended to neglect global self-concept (Marsh, 1990), instead of drawing on other than second-order factor models (but see Brunner et al., 2010) to model global self-concept. Thus, global self-concept remained an elusive construct.

A corresponding empirical question is how global self-concept, defined as the apex of the self-concept hierarchy, is related to the unidimensional measure of global self-esteem. Theoretically, the idea that global self-concept and global self-esteem are conceptually similar is well-accepted (Marsh & Craven, 2006; Marsh, Craven, & Martin, 2006). A second-order factor representing global self-concept showed very high correlations with global self-esteem (Marsh & Hattie, 1996). However, given the difficulties encountered in modeling the second-order factor presented above (i.e., low variances), it is unclear what these correlations mean on a

practical level. At the same time, researchers studying global self-esteem have expressed doubt about the relevance of domain-specific self-concepts for global self-esteem (Harris, Wetzel, Robins, Donnellan, & Trzesniewski, 2018; Orth & Robins, 2019). These implications have their foundations in longitudinal studies, which have suggested that domain-specific self-concepts have only a small amount of power for predicting global self-esteem (Marsh & O'Mara, 2008; Rentzsch, Wenzler, & Schütz, 2016; Trautwein, Lüdtke, Köller, & Baumert, 2006).

1.2 The Development of Global Self-Esteem

Along with understanding the nature of the global self, a great deal of theoretical work has focused on how the global self develops. From now on, I will focus on the framework of global self-esteem as a unidimensional, directly measured construct. Before investigating factors that can contribute to the development of self-esteem, it is of vital importance to review previous findings on the questions of whether and when self-esteem develops. When psychological researchers evaluate the development of a construct, they typically distinguish between *change*, which refers to shifts in mean levels across time, and *consistency*, which addresses changes in the relative standing of individuals within a group (Roberts & DelVecchio, 2000; Roberts, Walton, & Viechtbauer, 2006). In the following chapter, I aim to provide an overview of findings on the development (i.e., change and consistency) of self-esteem with a particular emphasis on the consistency of self-esteem.

After conducting studies using large and diverse samples, previous research concluded that the mean levels of self-esteem increase during childhood, peak in middle adulthood at about 50 to 60 years of age, and decline in old age (Orth, Trzesniewski, & Robins, 2010; Orth, Robins, & Widaman, 2012; Orth, Maes, & Schmitt, 2015; Orth, Erol, & Luciano, 2018). Using growth curve models, some of these studies have shown that an inverted U-shaped curve fits the lifespan data best. Some studies have focused specifically on the development of self-esteem in adolescence and young adulthood, yet their findings have been somewhat contradictory (Erol & Orth, 2011; Hutteman, Nestler, Wagner, Egloff, & Back, 2015; Morin, Maïano, Marsh, Nagengast, & Janosz, 2013; Soest, Wichstrøm, & Kvalem, 2016; Wagner, Lüdtke, Jonkmann, & Trautwein, 2013; Wagner, Lüdtke, Robitzsch, Göllner, & Trautwein, 2018). Whereas some studies have supported the finding that self-esteem increases in adolescence and young adulthood, a trend that is consistent with the maturity principle of personality development (Roberts & Wood, 2006), other studies have observed temporary declines, especially after the transition to secondary school, a finding that is in line with the disruption hypothesis during the phase of puberty (Soto & Tackett, 2015). Two recent meta-analyses averaged these findings and concluded that there was no change in self-esteem during adolescence (Orth et al., 2018; Scherrer & Preckel, 2019).

Along with investigations on change in self-esteem, multiple studies have targeted the consistency of self-esteem over time. Drawing on meta-analytical evidence, including longitudinal studies across the entire lifespan, the rank-order stability of global self-esteem in studies with an average time interval between assessments of about 3 years was low to medium during

childhood (r = .27 to .45), increased in adolescence (.46 to .61), peaked in adulthood (.49 to .72), and decreased during old age (.35 to.64; Trzesniewski, Donnellan, & Robins, 2003). On the basis of these findings, previous research concluded that the increases found in the consistency of self-esteem throughout adulthood are in line with the cumulative continuity principle of personality development (Caspi, Roberts, & Shiner, 2005; Roberts & DelVecchio, 2000), which suggests that personality becomes more and more stable with age. Besides using rank-order correlation coefficients, an alternative way to represent the consistency of constructs involves looking at the decomposition of a measure's variance over time (for details, see Chapter 1.5.2). As such, latent state-trait analyses indicated that a stable trait factor explained about 70% to 85% of the variance in self-esteem, and a state (residual) factor for each time point accounted for 15% to 30% of the variance (Donnellan, Kenny, Trzesniewski, Lucas, & Conger, 2012; Kuster & Orth, 2013). In a similar fashion, a 10-year longitudinal study that began when participants were young adolescents revealed that most of the variance in global self-esteem was accounted for by a stable and an autoregressive trait factor but that state (residual) variance still explained substantial (12% to 14%) amounts of variance (Wagner, Lüdtke, & Trautwein, 2016). Overall, these findings led researchers to conclude that self-esteem is a trait-like construct (Orth & Robins, 2014).

The reason that global self-esteem exhibits trait-like consistency might be because the nature of the construct is actually fixed or it might be due to self-selection and adaptive mechanisms (e.g., that individuals consistently seek information that confirms previous self-views, thereby contributing to the preservation of their self-esteem; Kuster & Orth, 2013). At the same time, however, it is surprising that no studies have questioned these findings, which have been treated as nearly axiomatic assumptions, given that the previously found stability could also have resulted from how self-esteem was measured. Interestingly, most of the longitudinal analyses described above have used self-esteem measures that were designed to *assess trait self-esteem*, by framing items in terms of very general, cross-situational, typical feelings and behaviors, most prominently the Rosenberg Self-Esteem Scale (Donnellan et al., 2015; Rosenberg, 1989). Even though Rosenberg did not make this alignment explicit, the way he designed his questionnaire clearly targeted trait-like self-esteem (e.g., "On the whole, I am satisfied with myself"). Yet, based on the assumptions set forth by Rosenberg (1986), Leary and Baumeister (2000), as well as other scholars (Heatherton & Polivy, 1991; Kernis, Cornell, Sun, Berry, & Harlow, 1993; Kernis, 2005), trait-like aspects are just one dimension of self-esteem, and state self-esteem is another relevant dimension. In fact, there are also other measures of self-esteem

that have particularly targeted these state-like aspects (Heatherton & Polivy, 1991; Ravens-Sieberer et al., 2001). Their application revealed that temporary fluctuations in state self-esteem can be a result of positive or negative feedback from others or a result of particular self-enhancement processes, which are often addressed in research on social psychology (Tesser, Millar, & Moore, 1988; Tesser, Crepaz, Collins, Cornell, & Beach, 2000).[3]

Given the theoretical importance of both trait and state self-esteem, the absence of studies that have simultaneously investigated and integrated findings on trait *and* state measures of self-esteem is surprising. However, this must be done so that a more granulated picture of the short- and long-term consistency of self-esteem can be developed. A central question would be whether and how rank-order stability varies across state and trait measures of self-esteem. In addition, it seems imperative to scrutinize the link between state and trait measures and the decomposition of state-trait variance.

[3] Along with the level of state self-esteem, fluctuations in and the fragility of state self-esteem itself have been the target of self-esteem research (e.g., Geukes et al., 2017; Kernis, 2005; Kernis, Cornell, Sun, Berry, & Harlow, 1993; Webster, Smith, Brunell, Paddock, & Nezlek, 2017).

1.3 Predictors and Consequences of Global Self-Esteem

Drawing on the overall conclusion that there is both change and consistency in global self-esteem, questions about predictors that contribute to differences in change as well as consequences that result from (stable) differences have emerged. In general, theoretical approaches to self-esteem have focused on global conditions of development in self-esteem (see Chapter 1.1.1). As part of the predictors, there has been general agreement that differences in the reflective appraisal and positive social ties from significant others precede differences in self-esteem (e.g., Cooley, 1902; Leary & Baumeister, 2000). James (1890/1963), Harter (1983, 1998), and Shavelson et al. (1976) additionally emphasized the role of domain-specific experiences in important areas of life. In terms of consequences, Rosenberg (1989) pointed in particular to the negative affective consequences of low self-esteem on anxiety and depression. Harter (1993) did not emphasize depression as a consequence of self-esteem but rather as a common factor along with self-esteem. Furthermore, Rosenberg (1989) also emphasized the social consequences of self-esteem. Predictors and consequences of self-esteem are of particular importance during adolescence and young adulthood because during this time, there is most potential for malleability of self-esteem (e.g., Harter, 1998; Rosenberg, 1986). More generally, regarding both predictors and consequences, global self-esteem can be expected to be related to its predictors and consequences on a comparable level of granularity (specificity matching principle; see Swann, Chang-Schneider, & Larsen McClarty, 2007), such as cross-situational experiences on the side of predictors or broad life outcomes on the side of consequences.

In order to study the predictors and consequences of self-esteem, it is useful to draw on a more global understanding of development. Multiple theoretical approaches have suggested that studying development across the lifespan, and in particular during adolescence and young adulthood, requires the integration of individual and environmental factors (e.g., Bronfenbrenner & Ceci, 1994; Lerner, 1998, 2006; Lerner, Lerner, Eye, Bowers, & Lewin-Bizan, 2011; Lerner, Lerner, Lewin-Bizan et al., 2011; Lewin, 1939; Magnusson & Stattin, 2006; Wagner, Orth, Bleidorn, Hopwood, & Kandler, in press). For example, in the theoretical framework of positive youth development, adolescents' development (e.g., the development of self-esteem) has been described as an interplay between "individual strengths" and "ecological assets" (Lerner, 1998, 2006; Lerner, Lerner, Eye, et al., 2011; Lerner, Lerner, Lewin-Bizan, et al., 2011). Individual strengths were described as individuals' cognitive, emotional, and behavioral resources. Ecological assets comprise the resources provided by the environment (i.e., families, schools, and communities). In previous research on self-esteem development, both individual

and environmental factors have been the subject of research on predictors *and* consequences (for an overview of predictors and consequences, see Orth & Robins, 2014, 2019). Interestingly, there has been large overlap between constructs that are considered predictors and those that are considered consequences of self-esteem (e.g., Harris & Orth, 2019; Sowislo & Orth, 2013). One reason for this may be that many relations between self-esteem and other constructs follow a transactional process, characterized by reciprocal associations between the two constructs (Sameroff, 2009; Swann et al., 2007). Another reason might be that the direction of the relations between self-esteem and these constructs has yet to be clarified. This has given rise to a more integrative consideration of the predictors and consequences of self-esteem. Therefore, this chapter provides a simultaneous overview of research on *individual predictors and consequences*, which is followed by a review of *environmental predictors and consequences*. In particular, I will review predictors and consequences that are relevant during adolescence and young adulthood. More specifically, I will place particular emphasis on (a) depressive symptoms as part of the *individual* predictors and consequences and (b) schools as social contexts as part of the *environmental* predictors and consequences.

1.3.1 Individual Predictors and Consequences

A variety of individual characteristics and experiences have been linked to self-esteem in both cross-sectional and longitudinal studies. For example, cross-sectional studies have suggested that males tend to have higher self-esteem than females, ethnic groups differ in their self-esteem (Bleidorn, Arslan et al., 2016; Kling, Hyde, Showers, & Buswell, 1999; Zuckerman, Li, & Hall, 2016), and the Big Five personality traits are related to differences in self-esteem, most prominently extraversion and neuroticism (Robins, Tracy, Trzesniewski, Potter, & Gosling, 2001). One rationale for explaining these relations could be derived from genotypic associations of self-esteem, which have been found, for example, in twin study designs (Bleidorn, Hufer, Kandler, Hopwood, & Riemann, 2018). In addition, stereotypical societal perceptions of individual characteristics (e.g., gender) may shape individuals' implicit or explicit perceptions of their self-esteem (Zuckerman et al., 2016).

Along with relatively fixed individual characteristics, there are multiple more malleable characteristics or experiences that have been linked to self-esteem. Previous longitudinal studies have supported the inference that stressful life events predict declines in individuals' self-esteem. For example, a serious disease, an accident, the loss of an important person, and break ups were found to be associated with declines in self-esteem, even after third variables were

accounted for (Orth & Luciano, 2015; Tetzner, Becker, & Baumert, 2016). In contrast to this, previous research observed no or small reciprocal relations between particular abilities (i.e., grades and achievement scores) or perceptions of particular abilities and global self-esteem (Baumeister, Campbell, Krueger, & Vohs, 2003; Harris et al., 2018; Marsh & O'Mara, 2008; Rentzsch et al., 2016; Trautwein et al., 2006). One reason for these findings is that domain-specific abilities and perceptions of them were too specific to be uniquely related to global self-esteem, a finding that would be in line with the specificity matching principle (Swann et al., 2007). More broadly, as described in Chapter 1.1.2, it is still unclear how different levels of the hierarchy of self-concept (i.e., domain-specific self-concepts and global self-concept) are related to global self-esteem. According to the specificity matching principle, it is likely that more global indicators are related to global self-esteem. Similarly, multiple studies have suggested that global indicators of success (or failure) in life such as delinquency and criminal behavior (Donnellan, Trzesniewski, Robins, Moffitt, & Caspi, 2005; Trzesniewski et al., 2006), economic prospects (Trzesniewski et al., 2006), or work-related outcomes such as job satisfaction, employment status, and salary (Kuster, Orth, & Meier, 2013) are prospectively linked to self-esteem. These studies suggest that success in life is a long-term consequence of global self-esteem rather than a predictor. A large body of research on self-esteem has pointed to the link between self-esteem and mental health problems (Orth & Robins, 2014). In particular, it has been argued that low self-esteem is reciprocally related to anxiety, well-being, physical health, and depression (e.g., Orth et al., 2012; Sowislo & Orth, 2013; Trzesniewski et al., 2006). In this field of research, the relation between self-esteem and depression has received the most theoretical and empirical interest, most likely because of the great importance that depression has for society (World Health Organization, 2008) in reference to high rates of lifetime prevalence (more than 15%; Kessler et al., 2005). In addition, depression is a particularly important concern during adolescence because this is the peak phase of first incidence (Kieling et al., 2019).

A Closer Look at Depressive Symptoms

There is broad agreement that depression is related to low self-esteem (e.g., American Psychiatric Association, 2013; Beck, 1967; Harter, 1993; Sowislo & Orth, 2013). Previously, it was hypothesized that depression and self-esteem represent opposite ends of a continuum that characterizes one construct (Harter, 1993; Watson, Suls, & Haig, 2002), yet this assumption has been largely ruled out empirically (Orth, Robins, & Roberts, 2008; Rieger, Göllner, Trautwein, & Roberts, 2016). Hence, a longstanding interest of researchers is to understand the

causal direction of the relation between depression and self-esteem. In this regard, previous clinical and personality researchers have primarily contrasted two theoretical ideas: On the one hand, depression has been presented as a predictor of low self-esteem, described in the framework of the *scar model* (Lewinsohn, Steinmetz, Larson, & Franklin, 1981). More specifically, the scar model states that depressive episodes leave enduring psychological scars on individuals. This can occur because depression leads to long-term difficulties in social functioning and in global attributions of the self, which can result in impaired self-esteem (Lewinsohn et al., 1981; Lewinsohn, Hoberman, & Rosenbaum, 1988; Shahar & Davidson, 2003). On the other hand, depression has been described as a consequence of low self-esteem, a process that has been outlined as the *vulnerability model* (Beck, 1967). The assumption underlying this model is that individuals with low self-esteem are at particular risk of exhibiting depression. In line with the diathesis-stress model in the cognitive theory of depression (Beck, 1967), low self-esteem represents a diathesis factor for depression. When it comes to stressful circumstances in life, this diathesis factor determines the direction taken at a crossroad: Whereas high self-esteem serves as a resource for resilience in coping with stressful life circumstances (Orth et al., 2008), low self-esteem leads to greater vulnerability to rejections and failures (Shahar & Davidson, 2003).

These two competing theories have been studied intensively in longitudinal studies. Across different age groups, gender distributions, and countries, the majority of studies have observed that the paths from self-esteem to depressive symptoms[4] have been higher than the opposite paths (Orth et al., 2008; Orth, Robins, & Meier, 2009; Orth & Robins, 2013; Rieger et al., 2016; Sowislo & Orth, 2013). Consequently, this field of research has agreed that the vulnerability model has empirical advantages over the scar model. However, it is important to note that in all of these studies, self-esteem was more stable over time than depressive symptoms. This is most likely the result of differences in the measurement of the two constructs. As described in Chapter 1.2, self-esteem has typically been measured as a trait (i.e., stable evaluations of self-worth). By contrast, depressive symptoms have been assessed with a more state-like measurement (i.e., to capture temporary or weekly feelings and thoughts). Another problem from previous studies is that all the studies were based on the same analytical approach, namely, between-person cross-lagged panel models. I identified only two very recent studies on the relation between self-esteem and depressive symptoms that applied other configurations

[4] These studies referred to depressive symptoms as a continuous variable, which is why from here on, I refer to depressive symptoms rather than depression (see also Orth et al., 2008).

of cross-lagged models with a stronger within-person focus. Both studies used meta-analytic methods across multiple data sets, and their results did not mirror the clear support for the vulnerability model that came from the traditional cross-lagged panel models (Masselink et al., 2018; Orth, Clark, Donnellan, & Robins, 2020). A closer look at the different methodological representations and their theoretical meanings is therefore needed. In sum, to date, research has supported the assumption that self-esteem is a vulnerability factor for depressive symptoms. However, there are multiple reasons for why this finding needs to be reconsidered with a greater focus on the measurement and modeling of the constructs.

1.3.2 Environmental Predictors and Consequences

Is the environment that surrounds adolescents relevant for their self-esteem? And are environmental factors themselves affected by individual differences in individuals' self-esteem? In order to gain a deeper understanding of these questions, I want to provide an overview of the findings on the relevance of the cultural and social environment and put particular emphasis on the school as a social environment. The cultural context has been described as an important environmental factor for individuals' self-related cognitions (e.g., Schmitt & Allik, 2005; Tafarodi & Swann, 1996). A recent comprehensive cross-sectional study found that cultures differed with respect to gender and age-related trajectories of self-esteem (Bleidorn, Arslan et al., 2016). These differences were partly moderated by differences in socioeconomic, demographic, and cultural-value indicators. Besides differences between cultures, there was support for shifts in self-esteem in the same culture over time and between cohorts in one study (Twenge, Carter, & Campbell, 2017), which had been hypothesized because of possible sociocultural changes toward higher self-regard. Yet, other studies did not find differences across cohorts (Erol & Orth, 2011; Orth et al., 2010; Orth et al., 2015). A more narrow cultural perspective was adopted by Bleidorn, Schönbrodt, et al. (2016), who observed higher self-esteem for individuals whose personalities (e.g., their emotional stability) and religiosity matched the characteristics of the city they lived in. Even though the effect sizes were small, the study could be an indicator that individuals who live around similar others perceive a higher sense of belonging, which confirms their self-esteem (Leary & Baumeister, 2000).

More generally speaking, social perspectives on the theoretical origins of self-esteem have strongly emphasized social interactions as a major environmental factor involved in self-esteem (Bowlby, 1969, 1973, 1980; Cooley, 1902; Deci & Ryan, 1985; Leary & Baumeister, 2000, see Chapter 1.1.1). Accordingly, individuals internalize social experiences and strive for

attachment and belonging in order to maintain and enhance a positive sense of the self. Although positive social relationships are primarily thought of as predictors of self-esteem, other theories have suggested that individuals' self-esteem itself shapes the subjective perception of social relations as well as the actual quality and quantity of relationships. For example, self-broadcasting theory suggests that individuals express their self-evaluations in their social behavior (Srivastava & Beer, 2005). Thus, individuals with high self-esteem might be more confident in the social context, and therefore, they potentially obtain more and better social ties. Overall, previous studies have suggested that the quality and quantity of social relationships predict changes in self-esteem (e.g., Gruenenfelder-Steiger, Harris, & Fend, 2016; Krauss, Orth, & Robins, 2019; Mund, Finn, Hagemeyer, Zimmermann, & Neyer, 2015; Orth, 2018; Reitz, Motti-Stefanidi, & Asendorpf, 2016; Wagner et al., 2013; but see Harris et al. 2015). At the same time, studies have less consistently supported effects that go in the opposite direction from self-esteem to social relationships (e.g., Gruenenfelder-Steiger et al., 2016; Orth et al., 2012; Schaffhuser, Wagner, Lüdtke, & Allemand, 2014). A recent meta-analysis aimed to bring together the two opposing directions of effects (Harris & Orth, 2019). They observed small but significant bidirectional effects between self-esteem and social relationships, whereas the effect from social relations to self-esteem was stronger than the opposite effect. This supported the assumptions made by the social perspectives on self-esteem, such as sociometer theory (Leary & Baumeister, 2000). Across this research field and the corresponding meta-analysis, the social context of families and peers were the focus of research. This is reasonable given that these are the most obvious and typically the closest social contacts in the lives of adolescents. However, other social contacts have gained surprisingly little attention. In particular, there is a paucity of research that has addressed social contexts in school as an important environment where self-esteem development might take place.

A Closer Look at Schools as Social Environments

In children and adolescents' lives, the school context represents an important setting that surrounds and affects them every day. In school, they develop academically and personally, which makes it likely that self-evaluations manifest in this environment. According to stage-environment-fit theory by Eccles and Midgley (1989), the school environment is crucial for adolescents' socioemotional development because of the possible mismatch that can occur between students' changing psychological, motivational, and emotional attributes and the experiences and characteristics of their schools. A mismatch between students' development and

their school environment has been shown to appear in particular after the transition to secondary school (e.g., Roeser, Eccles, & Sameroff, 2000), and previous findings have suggested that this tends to be accompanied by declines in self-esteem (Wigfield & Eccles, 1994).

In-school social interactions take place with classmates and teachers. Previous research has indicated that perceptions of being popular in the classroom are associated with changes in self-esteem (e.g., Gruenenfelder-Steiger et al., 2016; Reitz et al., 2016; Wagner et al., 2018). Moreover, social comparisons with classmates were strongly linked to self-perceptions. However, this was particularly true for domain-specific academic self-concepts but not necessarily for global self-esteem (e.g., Marsh, 1987a; Marsh & Hau, 2003; Marsh, Trautwein, Lüdtke, & Köller, 2008). Social interactions with teachers, however, have not been given much attention in previous research on self-esteem. This is surprising given that students often strive for emotional security and acceptance from teachers, which is why teachers have been described as the most important nonfamiliar people in students' lives (Kesner, 2000). This influence might reach beyond support for academic issues because teachers can also act as confidants for personal problems and can provide support for behavioral issues (Ryan, Stiller, & Lynch, 1994). Accordingly, students' relationships with teachers might affect students' very general feelings about the self, ranging from affective feelings such as self-confidence and coping strategies to cognitive-motivational feelings such as competence and control. At the same time and in line with self-broadcasting theory (Srivastava & Beer, 2005), teachers' behaviors might also be influenced by students' self-esteem in the classroom because students compete for attention from their teachers. Moreover, from a risk regulation perspective (Murray, Holmes, & Griffin, 2000; Murray, Holmes, & Collins, 2006), high self-esteem students might at least *perceive* better relationships with their teachers because students transfer their feelings about the self to feelings about others. A few empirical studies have suggested that positive student-teacher relationships are associated with more positive adjustment, such as self-esteem (Aldrup, Klusmann, Lüdtke, Göllner, & Trautwein, 2018; Ryan et al., 1994). However, previous studies have not analyzed this idea in a comprehensive, reciprocal, longitudinal design.

Parallel to the findings on self-esteem and depressive symptoms, previous research on the reciprocal relations between social relationships and self-esteem has largely been based on traditional (cross-lagged) regression models that were aimed at investigating *between-person* effects (Harris & Orth, 2019). These results indicated how individual differences in social relationships were associated with individual differences in self-esteem. One recent study analyzed *within-person* relations between family environment and self-esteem (Krauss et al., 2019)

and found a similar but less consistent pattern of within- compared between-person effects (i.e., family environment predicted self-esteem). In addition, the study described convergence problems in the within-person analyses and pointed to conceptual problems that can occur in within-person analyses when constructs reveal high rank-order stability. Hence, a closer consideration of how these conceptual ideas are related to modeling procedures is needed. In particular, no research has targeted these considerations for the relation between student-teacher relationships and self-esteem.

1.4 Recapping the Questions that have Emerged about Self-Esteem

In the previous chapter, I gave an overview of the theoretical and empirical background regarding the origins and different conceptualizations of the global self, the development of global self-esteem, as well as the predictors and consequences of global self-esteem. Theoretical approaches to self-esteem have highlighted the role that social interactions play in determining self-esteem (e.g., Cooley, 1902; James, 1890/1963; Leary & Baumeister, 2000) and the particular importance of the phase of adolescence for the development of self-esteem (e.g., Harter, 1998; Rosenberg, 1986). Whereas most of the research on the global self has focused on a unidimensional conceptualization of global self-esteem (Rosenberg, 1989), some approaches have pointed to a multidimensional, hierarchical nature with global self-concept at the apex of this hierarchy (Harter, 1998; James, 1890/1963; Shavelson et al., 1976). Longitudinal data over the lifespan have indicated that mean levels of self-esteem tend to be highest during adulthood (Orth et al., 2018). At the same time, findings during adolescence tend to be somewhat contradictory; whereas some found increases, others observed decreases in self-esteem during adolescence (for a meta-analysis, see Orth et al., 2018). Moreover, longitudinal studies have observed increases in the consistency of self-esteem throughout adulthood, which led to conclusions about the trait-like nature of self-esteem (Trzesniewski et al., 2003). According to previous findings, there were both individual and environmental predictors and consequences of self-esteem. Many constructs are considered to function as both predictors *and* consequences. In terms of *individual* predictors and consequences, aspects of mental health have garnered particular interest (Sowislo & Orth, 2013), whereas *environmental* predictors and consequences in particular have tended to focus on social relationships (Harris & Orth, 2019). Despite a long history and a comprehensive research field, there were, and still are, emerging questions, which have yet to be answered. Specifically, I want to highlight three areas of interest.

First, along with a focus on unidimensional global self-esteem, previous research pointed to global self-concept as the apex of a multidimensional hierarchy (Shavelson et al., 1976). Yet, previous research primarily modeled global self-concept by using a second-order factor approach, but it did not fit the data well. These findings led to the neglect of global self-concept, rather than to a tendency to more strongly consider the theoretical fit of second-order factors in the self-concept hierarchy. In order to gain a better understanding of global self-concept, a closer alignment of theoretical and methodological considerations is needed. In addition, given

the previous challenges involved in modeling global self-concept, it is largely unclear how global self-concept as the top of the hierarchy is linked to global self-esteem.

Second, unidimensional global self-esteem was observed to be largely consistent over time as indicated by rank-order stabilities and the decomposition of state-trait variance. These findings led researchers to conclude that self-esteem has a trait-like nature (Orth & Robins, 2014). At the same time, these inferences were based on measures that were actually designed to target only trait self-esteem. However, in stark contrast, theoretical approaches have focused on both state and trait self-esteem (Leary & Baumeister, 2000). There is a lack of studies that have simultaneously addressed both state and trait measures, let alone studies that have merged how the conceptualization of state and trait measures is related to the modeling of proportions of state and trait variance. In addition, it is unclear how the so far predominantly trait-like measurement of self-esteem has influenced research on its relations with other constructs, such as depressive symptoms.

Third, there is a great deal of theoretical and empirical evidence that individuals' self-esteem gets shaped in the social context through interactions with significant others (Cooley, 1902; Harris et al., 2018; Leary & Baumeister, 2000). At the same time, self-esteem might also contribute to the quality and quantity of social interactions (Srivastava & Beer, 2005). Much attention has been attributed to the role of parents and peers, whereas other social contexts have gained surprisingly little interest. Specifically, the role of teachers has not been addressed sufficiently even though teachers have been described as the most important nonfamiliar attachment figures (Kesner, 2000). Reciprocal relations between social relationships and self-esteem have typically been investigated via traditional cross-lagged panel models to target between-person effects, whereas within-person analyses are still rare.

Interestingly, all of these areas of interest and the resulting research questions are fundamentally connected to their methodological implementation. Therefore, it is particularly important to obtain a more sophisticated understanding of the assumptions and consequences of different methodological approaches for addressing these research questions, which could, ultimately, lead to synergistic effects that can be applied to answer these emerging questions about self-esteem.

1.5 Integrating Theoretical and Methodological Considerations

In an attempt to integrate theory and methods, researchers need to scrutinize the settings, measures, and analytical models that allow them to draw inferences about the theoretical phenomena of interest (Cook & Campbell, 1979). In this regard, psychological researchers often refer to the concept of *construct validity* as the "degree to which inferences are warranted from the observed persons, settings, and cause and effect operations […] to the constructs that these instances might represent" (Shadish, Cook, & Campbell, 2002, p. 38). Shadish et al. (2002) emphasized that threats to construct validity (e.g., an inadequate or confounding explication of the construct) can jeopardize the entire research undertaking, given that a high degree of construct validity is an important prerequisite for drawing conclusions that can refer back to the theoretical phenomena.

Originally, construct validity was evaluated in the context of psychological test scores embedded in the question of whether a test measures what it claims to measure (Cronbach & Meehl, 1955). From this perspective, researchers assess whether the pattern of relations that a test shows with other constructs reflects the theoretical phenomena of interest (Campbell & Fiske, 1959; Cronbach & Meehl, 1955). In reference to the definition of construct validity proposed by Shadish et al. (2002), construct validity should be considered from a more comprehensive perspective on *all* operational aspects of a research question (not only for particular tests and measures). Moreover, despite never reaching a state of perfect construct validity (Borsboom, 2006; Cronbach & Meehl, 1955; Messick, 1988) and rather than seeing construct validity as a characteristic of test scores, this concept should remind researchers to constantly evaluate the degree to which theoretical assumptions and their empirical implementations are appropriate and justified (e.g., Borsboom, Mellenbergh, & van Heerden, 2004; Shadish et al., 2002; Smith, 2005).[5] Thus, in contrast to earlier assumptions about construct validity (Campbell & Fiske, 1959; Cronbach & Meehl, 1955), more recent deliberations have emphasized a stronger theoretical focus, which begins long before patterns of correlations are assessed. Instead, it begins when the construct is defined and the choice of measures and methods are scrutinized in reference to the theoretical definition (Borsboom et al., 2004; Smith, 2005). According to Marsh, Martin, and Hau (2006) and Marsh and Hau (2007), a construct validation

[5] Borsboom et al. (2004) and Borsboom, Cramer, Kievit, Zand Scholten, and Franic (2009) suggested that researchers should refrain from using the term "construct validity" in reference to the problems involved in previous conceptions of construct validity as a characteristic of *test scores*. By contrast, they suggested that researchers should embrace the more global term "(test) validity" in reference to the (theoretical) properties of *tests*.

perspective includes taking into consideration the interwoven and interdependent relations between theory, measurement, empirical research, and practice.

Besides constructing specific measures to assess a construct, another important part of this construct validity approach includes addressing the selection of appropriate statistical models that depict the theoretical research question and test it statistically (Borsboom et al., 2004; Marsh & Hau, 2007). In psychological research, unobserved theoretical constructs can be modeled by using observed measures within the framework of latent variable models (Borsboom et al., 2003). Latent variable models offer a flexible set of tools that can be applied to analyze complex multivariate, longitudinal, and multilevel theoretical phenomena. Yet, with the rise of latent variable models, Marsh and Hau (2007, p. 155) identified "an ever widening gap between 'state-of the-art' methodological and statistical techniques […] and the actual skill level of many applied researchers." Thus, previous research has often suffered from a lack of an actual integration of theory and methods because scientific journals and training programs have focused on *either* theoretical *or* methodological questions in psychology (Borsboom, 2006; Marsh & Hau, 2007). Yet, particularly when it comes to nonexperimental designs or complex data structures, sophisticated statistical solutions are necessary for representing the theoretical phenomena of interest appropriately. Accordingly, Herbert Marsh and his colleagues proposed and coined a research agenda "that is at the cutting edge of both latest methodological developments and substantive issues – *methodological-substantive synergies*" (Marsh & Hau, 2007, p. 168). In response, in a very extensive series of papers, the group of researchers working with Herbert Marsh applied sophisticated methodological advances to address substantive research questions (e.g., Marsh, Lüdtke, Trautwein, & Morin, 2009; Marsh, Lüdtke, Robitzsch et al., 2009; Marsh, Muthén et al., 2009; Marsh, Liem, Martin, Morin, & Nagengast, 2011; Marsh, Nagengast, & Morin, 2013; Marsh, Kuyper, Morin, Parker, & Seaton, 2014; Marsh & Scalas, 2018; Morin, Maïano, Marsh, Janosz, & Nagengast, 2011; Morin et al., 2013; Morin, Arens, & Marsh, 2016; Nagengast & Marsh, 2011; Nagengast et al., 2011). More generally, past psychological research has demonstrated that methodological innovations were often generated from substantive theoretical questions, and at the same time, methodological developments substantially shaped scientific understanding and inspired further theoretical advances (Greenwald, 2012; Marsh & Hau, 2007).

Most often, there is not only one methodological approach that provides the only solution to a theoretical question, but it is fruitful to derive multiple perspectives (e.g., multiple methods to assess or analyze data) and compare their consequences (Marsh, Martin et al., 2006; Marsh

& Hau, 2007). Multiple operationalizations of a research question can help to reduce the bias that can occur when only one specific methodological approach is applied, which can result in an underrepresentation or confounding of the phenomena of interest (Marsh, Martin et al., 2006; Marsh & Hau, 2007; Shadish et al., 2002). Furthermore, instead of relying on one method to analyze a research question, multiple methodological perspectives can enhance the transparency of the research, given that they reduce the risk for a selective results-driven presentation of research findings (Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016).

In reference to the three emerging areas of interest presented above (conceptualization of higher order global self-concept, state and trait self-esteem, link between social relationships with teachers and self-esteem), the present dissertation attempts to merge theoretical questions with different methodological implementations. Therefore, I will be zooming in on methodological representations of *higher order constructs*, *states and traits*, and *reciprocal relations*.

1.5.1 Higher Order Constructs

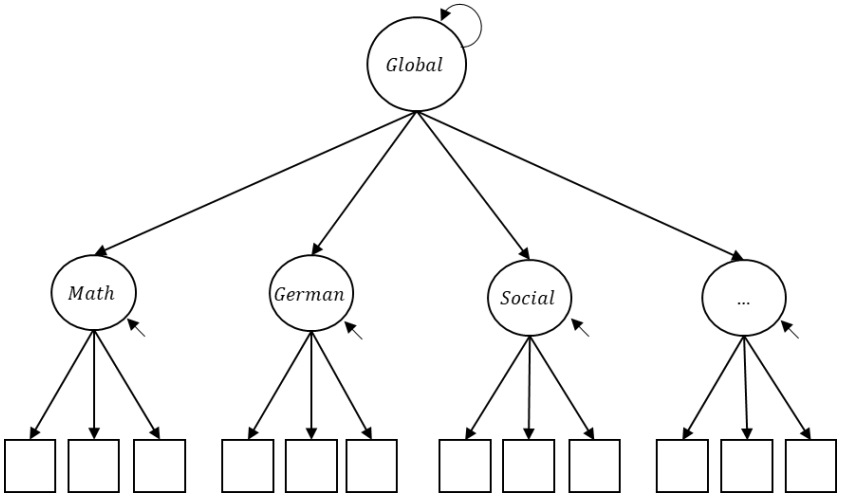
The first emerging question identified in this dissertation addresses the conceptualization of global self-concept as the apex of a multidimensional hierarchy of self-concept (Shavelson et al., 1976). Along with hierarchical self-concept, there is a variety of psychological constructs that can be conceptualized as hierarchical in nature. Among the most prominent examples are intelligence (e.g., Cattell, 1940, 1963; Gustafsson, 1984; Vernon, 1950) and well-being (e.g., Chen, West, & Sousa, 2006; Gallagher, Lopez, & Preacher, 2009). Hierarchical constructs encompass subconstructs at different levels of the hierarchy. For example, in hierarchical self-concept, there is global self-concept at the apex of the hierarchy and more domain-specific self-concepts at lower levels of the hierarchy (Marsh & Shavelson, 1985; Shavelson et al., 1976). Neither domain-specific self-concepts nor global self-concept can be observed directly but rather have to be operationalized through observable measures and further modeling approaches (Brunner, Nagy, & Wilhelm, 2012). In self-concept research, there are well-established self-report inventories that are designed to measure domain-specific self-concepts, which can be assessed in order to make inferences to the theoretical constructs on the lower level of the hierarchy of self-concept (Marsh & O'Neill, 1984). When it comes to higher order levels of the hierarchy, there have been diverging definitions on how to operationalize these more global constructs. On the one hand, global self-esteem has been measured directly, such as with the Rosenberg Self-Esteem Scale (Rosenberg, 1965, 1989). On the other hand, global self-concept has been modeled by using lower order self-concepts (Marsh & Shavelson, 1985). When mod-

eling global self-concept using lower order self-concepts, the crucial question is how this modeling should be implemented. This question needs to be grounded in theoretical assumptions about the conceptualization of global self-concept. Theoretically, on the one hand, global self-concept could be conceptualized via a top-down process by which global self-concept affects the lower order self-concepts, but on the other hand, it could be a bottom-up process by which lower order self-concepts form global self-concept. These two theoretical ideas can be represented through different methodological implementations.

Methodological implementations. A latent variable model is a strong tool that can be used to operationalize unobserved higher order constructs (e.g., global self-concept) on the basis of individuals' standing on lower order constructs (e.g., domain-specific self-concepts). Based on the underlying theoretical mechanisms, there are two broad types of latent variable models (Bollen & Lennox, 1991), which can be applied to model relations between indicators and first-order latent variables, but also, as in the present case, between higher order and lower order latent variables (Diamantopoulos, Riefler, & Roth, 2008): In *reflective models*, the underlying higher order construct causes differences in the indicators (here: lower order self-concepts), which is why these indicators are often described as effect indicators (Blalock, 1964). These models draw on classical test theory, which conceptualizes indicators as determined by the latent variable (here: the higher order construct) and an error term (Lord & Novick, 1968). As such, the latent variable itself is free from measurement error. Reflective models call for high correlations between the indicators because they should represent a unidimensional construct (Bollen & Lennox, 1991). The most prominent implementation of this idea on the level of higher order constructs would be a second-order factor model (see Figure 1, Panel A), which explains the common variance between the first-order factors.[6] Reflective models are the most common models in psychological research. Yet, it is unclear whether their preponderance is driven by the theoretical fit between reflective models and psychological constructs, or rather by their popularity and simple implementation (Rhemtulla, van Bork, & Borsboom, 2015, 2019). Thus, it is possible that the easy availability of the respective software and the widespread understanding of reflective measurement models as the state-of the art method have fundamentally shaped the rise of reflective measurement models (Rhemtulla et al., 2019).

[6] Along with second-order factors, there are also other configurations of reflective structural models that are aimed at depicting hierarchical constructs, such as bifactor models (e.g., Brunner et al., 2010; Brunner et al., 2012; Gustafsson & Balke, 1993).

A) Second-order factor



B) Model-based latent composite score

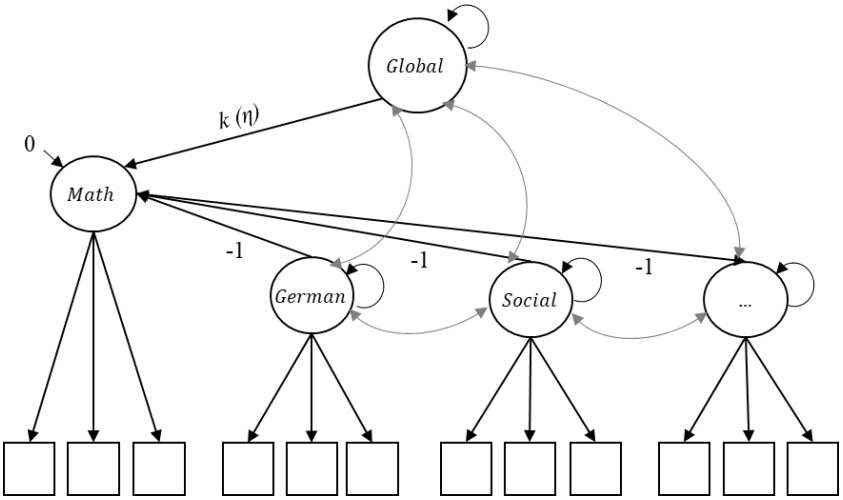


Figure 1. Simplified graphical representation of the second-order factor and the model-based latent composite score for global self-concept. Residual variances of the indicators are not displayed.

By contrast, in *formative models*, the indicators (here: lower order self-concepts) are conceptualized as causes of the higher order construct. Thus, the construct is formed by the indicators, and this conceptualization does not require substantial correlations between the in-

dicators (Bollen & Lennox, 1991). In these models, the measures are described as causal, formative, or composite indicators[7] (Blalock, 1964; Bollen & Lennox, 1991; but see Bollen & Bauldry, 2011). Yet, challenges in the implementation of different configurations of formative models (e.g., model identification or handling of missing data) might have resulted in a steady turning away from formative models (Bollen & Diamantopoulos, 2017). Surprisingly, conceptual questions about the nature of the underlying theoretical construct have often played a minor role in the selection of the model (Edwards & Bagozzi, 2000; Rhemtulla et al., 2019). One reason for this status could be that an empirical implementation using sophisticated latent variable models with a formative modeling approach had been missing. A new methodological development could provide this missing piece (Rose, Wagner, Mayer, & Nagengast, 2019; for details, see Study 1 in this dissertation): Inspired by latent change score models, the model-based latent composite score (see Figure 1, Panel B) enables researchers to build composite scores that are free from measurement error and to incorporate full information maximum likelihood estimation. The latent composite score is also comparably easy to implement in conventional statistical software (Rose et al., 2019). In addition, in contrast to traditional formative measurement models, the model-based latent composite score does not require an outcome variable in order to be identified (Bollen & Davis, 2009; Bollen & Bauldry, 2011).

Integrating methods and theoretical assumptions about global self-concept. In previous research on the hierarchy of self-concept, global self-concept was typically modeled by using a reflective model (i.e., second-order factor; e.g., Marsh & Shavelson, 1985; Marsh, 1987b). However, this revealed empirical problems (e.g., small variances for global self-concept) because the lower order self-concepts did not show substantial correlations. Interestingly, the idea of applying reflective models to model global self-concept was primarily inspired by the rise of these models in the field of intellectual abilities (Shavelson et al., 1976; Vernon, 1950). A reflective model on the hierarchy of self-concept would assume a top-down conceptualization in that global self-concept causally affects lower order self-concepts. At the same time, theoretical ideas about the relations between global self-concept and lower order self-concepts were more aligned with the process by which global self-concept is formed

[7] Whereas some papers use the terms causal, formative, or composite indicators interchangeably (e.g., Bollen & Lennox, 1991), more recent publications distinguish between causal and composite indicators as different classes of constructs and refrain from the term formative indicators (Bollen & Bauldry, 2011). However, in the present study, I refer to composites as part of the conceptual idea of formative models and, therefore, I do not further distinguish the terms composite and formative indicators (respectively composite and formative models).

(Shavelson et al., 1976). There has been little research on a formative perspective on the hierarchy of self-concept. Interestingly, the few existing studies go back to the time before the Shavelson model was introduced. However, back then, self-concept items were barely theoretically defined (e.g., Coopersmith, 1967; Fitts, 1965), and sophisticated formative modeling approaches that could account for missing data or measurement error were missing. A new empirical method can now provide a sophisticated approach that can be used to implement the process by which global self-concept is formed using a model-based latent composite score. Subsequently, an empirical comparison of the consequences of a reflective versus a formative representation (see Figure 1) is needed in order to improve the understanding of the elusive construct of global self-concept as well as the differences resulting from such implementations of higher order constructs.

1.5.2 States and Traits

The second emerging question outlined above (see Chapter 1.4) is the question of how stable self-esteem is, given that previous research used an almost axiomatic approach to measure trait-like self-esteem and subsequently observed large amounts of trait variance (Orth & Robins, 2014; Trzesniewski et al., 2003). At the same time, some scholars have emphasized the role of state self-esteem (e.g., Heatherton & Polivy, 1991; Leary & Baumeister, 2000). More generally, the differentiation of states and traits is an important concept across psychological research, which classifies constructs regarding their consistency over time (Eysenck, 1983). Traits are defined as "relatively enduring psychological characteristics that influence people's thoughts, feelings and behaviors" (Nezlek, 2007, p. 791). For many decades, the trait concept was the driving force in personality research (McCrae & Costa, 2008), but this focus shifted to some extent when personality researchers began to observe substantial changes in personality constructs across the lifespan (e.g., Nezlek, 2007; Roberts & DelVecchio, 2000; Roberts et al., 2006; Roberts, 2009). States are described as thoughts, feelings, and behaviors that change across time and situations (Nezlek, 2007). The state perspective has mainly guided research on affective constructs such as anxiety and mood (Spielberger, 1966). Yet, states have conquered a broader range of psychological disciplines, such as educational or personality research (e.g., Nezlek, 2007; Rieger et al., 2017; Roberts et al., 2006; Roberts, 2009, 2018). Even if, on the basis of their theoretical definitions, some constructs might be better suited as traits and others as states, it has been argued that most constructs need to be considered at different levels on a state-trait continuum (Epstein & O'Brien, 1985; Fleeson & Jayawickreme, 2015;

Geiser, Götz, Preckel, & Freund, 2017; Hertzog & Nesselroade, 1987; Rieger et al., 2017; Roberts & Wood, 2006; Roberts, 2018; Wrzus & Roberts, 2017).

Methodological implementations. Despite a consensus on the assumption that traits are more enduring and states are more situational, different objectives across empirical studies as well as different assumptions about the nature of states and traits have led to different empirical operationalizations of states and traits. These operationalizations have come primarily from two perspectives that have mostly functioned independently:

The first operationalization addresses state and trait *measures* of a construct. This means that researchers implicitly or explicitly place their construct of interest along the state-trait continuum by choosing an appropriate time frame (e.g., "During the last hour…"; "During the last two weeks…"; "In general…") with respect to which participants are asked to rate their answers. Accordingly, the time frames used in a questionnaire can be considered implementable and observable translations of states and traits. Therefore, the time frames need to be evaluated critically with regard to their degree of construct validity when referring back to the theoretical construct and the corresponding research question. The pioneer measurement-based differentiation of traits and states has been made in reference to anxiety (Spielberger, 1966), followed by other affective constructs, such as positive and negative affect or depression (Spielberger, 1995; Watson, Clark, & Tellegen, 1988). Yet, with regard to most other psychological constructs, researchers have tended to stick to only one specific time frame across different studies. However, a re-evaluation of whether the time frame matched the underlying theoretical construct of interest in different research questions has rarely been made. For example, when interested in short-term shifts in a construct, the time frame of "In general…" might not target the theoretical phenomenon of interest. There has been no universal answer to the question of which time frame is appropriate for assessing states and traits. One reason for this is the complexity of this question, given that it needs to be addressed in the context of a specific construct, a specific design, and a specific research question. For example, when setting different constructs in relation to each other, an important consideration would be whether the time frames of different constructs need to be comparable in order to rule out confounds that are due to the time frame.

The second operationalization of states and traits is based on a decomposition of the variance of observed measures into *trait and state (residual) variance* and an error term (Geiser et al., 2017; Kenny & Zautra, 1995, 2001; Steyer, Ferring, & Schmitt, 1992; Steyer, Schmitt, &

Eid, 1999), regardless of the time frames used to assess the constructs. Results on the decomposition of variance have been applied to address the stability of the measures over time but have also led researchers to draw conclusions about the state- or trait-like nature of a construct. Among the most prominent methodological approaches that have been applied to decompose variance over time is latent state-trait (LST) theory (Steyer et al., 1992; Steyer et al., 1999; see Figure 2) as an extension of classical test theory. According to LST theory, a psychological state is a compound of multiple aspects that led to the manifestation of the state observation. These aspects include stable characteristics of a person (traits), features of the situations, as well as the interaction between the person and the situation (Steyer et al., 1999). LST theory is based on the assumption that "measurement does not take place in a situational vacuum" (Steyer et al., 1999, p. 392) but always includes both characteristics of the person and characteristics of the situation. LST theory is in line with classical test theory, which proposes that any observed measure is a function of a true score variable and an error variable (Lord & Novick, 1968; see also the reflective model in Chapter 1.5.1). In LST theory, the true score variable is described as the *latent state variable*. Given multiple time points and therefore multiple latent state variables, the variance of the latent state variables can be further decomposed into trait variance, which is specific to the person, and latent state residual variance, which is due to the situation and the specific person in a situation (i.e., occasion-specific effects). Following the definition from LST theory, latent state residuals are uncorrelated with the trait factor and the error component.

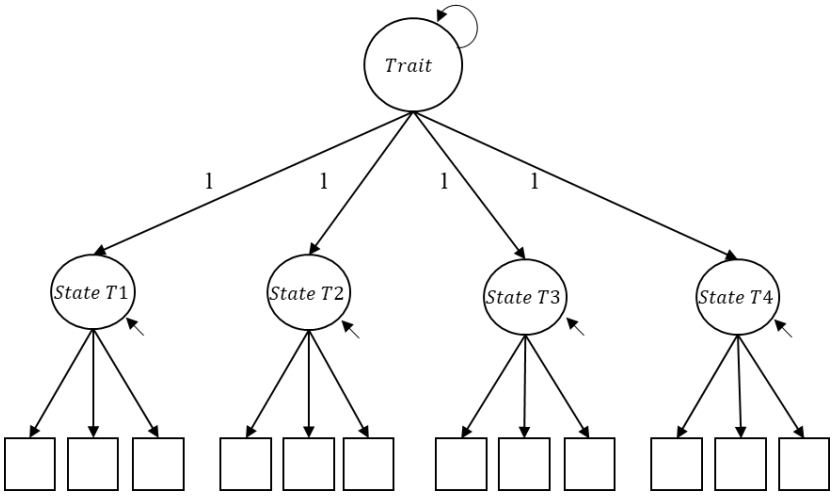


Figure 2. Simplified graphical representation of the latent state-trait model. The residual variances of the indicators are not displayed.

Taken together, research studies that have applied these two approaches to capturing the consistency of constructs have used the same terms (i.e., states and traits), yet they have operated fairly independently. Given that, in the past, researchers have often used the terms states and traits without further specifying whether they meant the measurement approach or the variance-proportion approach, this might have produced inaccuracies and confusion among researchers. The approaches are not opposed to each other; rather, they must be interwoven, given that both are aimed at classifying constructs on the basis of their consistency. More specifically, if a construct is measured with a trait time frame, it is likely to obtain larger amounts of trait variance than a construct with a state time frame. Parallel to the present dissertation, a recently published paper asked how state and trait measures of anxiety are related to the proportions of state (residual) and trait variance (Lance, Christie, & Williamson, 2019). They observed that both state and trait *measures* consisted of a majority of trait *variance*, yet state measures revealed (in total) less trait variance and more state (residual) variance than trait measures. Except for this first attempt to merge state and trait measures with the decomposition of state-trait variance, there has been a large gap between these two streams of research, and they have yet to be interlinked systematically.

Integrating methods and theoretical assumptions on state and trait self-esteem. Theoretically, self-esteem has mostly been considered a trait-like construct (i.e., a relatively enduring individual characteristic), yet there has also been research emphasizing state aspects of self-esteem (Donnellan et al., 2012; Heatherton & Polivy, 1991; Wagner et al., 2016). Previous research has usually applied a measure that incorporated a trait time frame (Rosenberg, 1989). There are less frequently used self-esteem measures that address more state-like approximations of self-esteem (Heatherton & Polivy, 1991; Ravens-Sieberer et al., 2001), but the consequences of applying state and trait self-esteem measures have not been analyzed systematically, and there has certainly not been systematic variation in bivariate research questions. Regarding the decomposition of state-trait variance, previous self-esteem research has indicated that the majority of variance in global self-esteem could be explained by a latent trait, whereas the latent state residual has been found to explain smaller but still substantial amounts of the variance (Donnellan et al., 2012; Wagner et al., 2016). However, these analyses have almost exclusively been based on trait measures of self-esteem and never included systematic variations in the measurement of state and trait self-esteem. One possible way to approach this research gap is to assess both state and trait measures of self-esteem and experimentally test their consequences for the decomposition of state-trait variance. Table 1 depicts four exemplary items

from the Rosenberg Self-Esteem Scale (RSE; Rosenberg, 1989). In the original version, the items were framed with a trait-like time frame ("In general…"). In addition, it would be possible to apply a different, more state-like time frame in order to construct a state measure of self-esteem. An experimental, longitudinal application of variations in the time frames could enrich the understanding of the interwoven nature between state-trait measures and state-trait variance components of global self-esteem.

Table 1
Sample Items from the Trait and State Versions of the Rosenberg Self-Esteem Scale (RSE)

Self-esteem trait	Self-esteem state
In general…	**During the last 2 weeks ...**
...I am satisfied with myself.	...I was satisfied with myself.
...I think I am no good at all. (R)	...I often thought I was no good. (R)
...I feel that I have a number of good qualities.	...I felt I had a number of good qualities.
...I am able to do things as well as most other people.	...I was able to do things as well as most other people.

Note. Trait self-esteem items stem from Rosenberg (1989). Items reproduced with permission from Wesleyan University Press. For a complete list of the items, see the Supplemental Material from Study 2 in the present dissertation.

1.5.3 Reciprocal Relations

Considerable research interest has been attributed to the reciprocal relations between self-esteem and individual variables, such as depressive symptoms (Sowislo & Orth, 2013), as well as environmental variables, such as social relationships (Harris & Orth, 2019). Studying prospective relations between different constructs over time is a central concern that has driven psychological research (Orth et al., 2020; Usami, Murayama, & Hamaker, 2019). Traditionally, cross-lagged panel models (CLPMs) have been applied to answer such research questions. However, recently, a debate on between- and within-person effects stimulated the creation of a variety of other analytical models for addressing reciprocal relations (for an overview of these models, see Orth et al., 2020; Usami, Murayama, et al., 2019; Usami, Todo, & Murayama, 2019; Zyphur, Allison et al., 2019; Zyphur, Voelkle et al., 2019). In a nutshell, these models differ in their assumptions about the need to explicitly model enduring between-person differences and trajectories over time. These technical variations come along with relevant differences in the interpretation of the paths of interest. Even though conceptual differences should

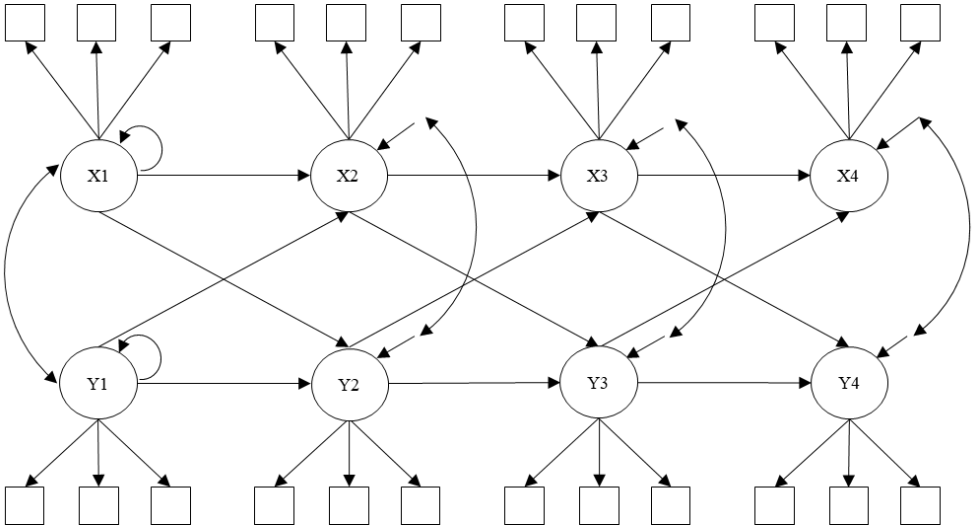
be the major concern of researchers when choosing one of these models for their research (Orth et al., 2020), many previous applications of and discussions about the models have suffered from justifications that were based on alleged empirical reasons, common standards, ideologies, or battles of power. This poses a clear threat to construct validity, given that the methodological implementation should be the most adequate translation of the theoretical research question. In the present dissertation, I want to narrow the perspective two three configurations of models to address reciprocal relations.

Methodological implementations. The most widespread analytical model that has been applied to analyze reciprocal relations over time is the cross-lagged panel model (CLPM; Biesanz, 2012; see Figure 3, Panel A). The CLPM indicates whether individual differences in one construct predict individual differences in the other construct when controlling for previous individual differences in the second construct. Hence, with the help of cross-lagged panel models, it is possible to address questions about prospective relations of overall individual differences in constructs (e.g., individual differences in student-teacher relationships and self-esteem; for an example, see Table 2). Despite the popularity of the CLPM, Hamaker, Kuiper, and Grasman (2015) set up an influential critique against traditional cross-lagged panel models—a critique widely received and adopted by the scientific community. According to Hamaker et al. (2015), a drawback of the CLPM is that it controls only for year-to-year stability but not for enduring "trait-like" differences over time (for a previous critique, see Rogosa, 1980). In line with the framework of multilevel models for longitudinal data (i.e., occasions nested within individuals), Hamaker et al. argued that the CLPM fails to disentangle the within-person level from the between-person level, which can result in fallacies regarding the effect on the within-person level. Therefore, it has been argued that researchers should control for stable between-person differences by modeling a random intercept for each construct (Cole, Martin, & Steiger, 2005; Hamaker et al., 2015; Usami, Murayama et al., 2019).

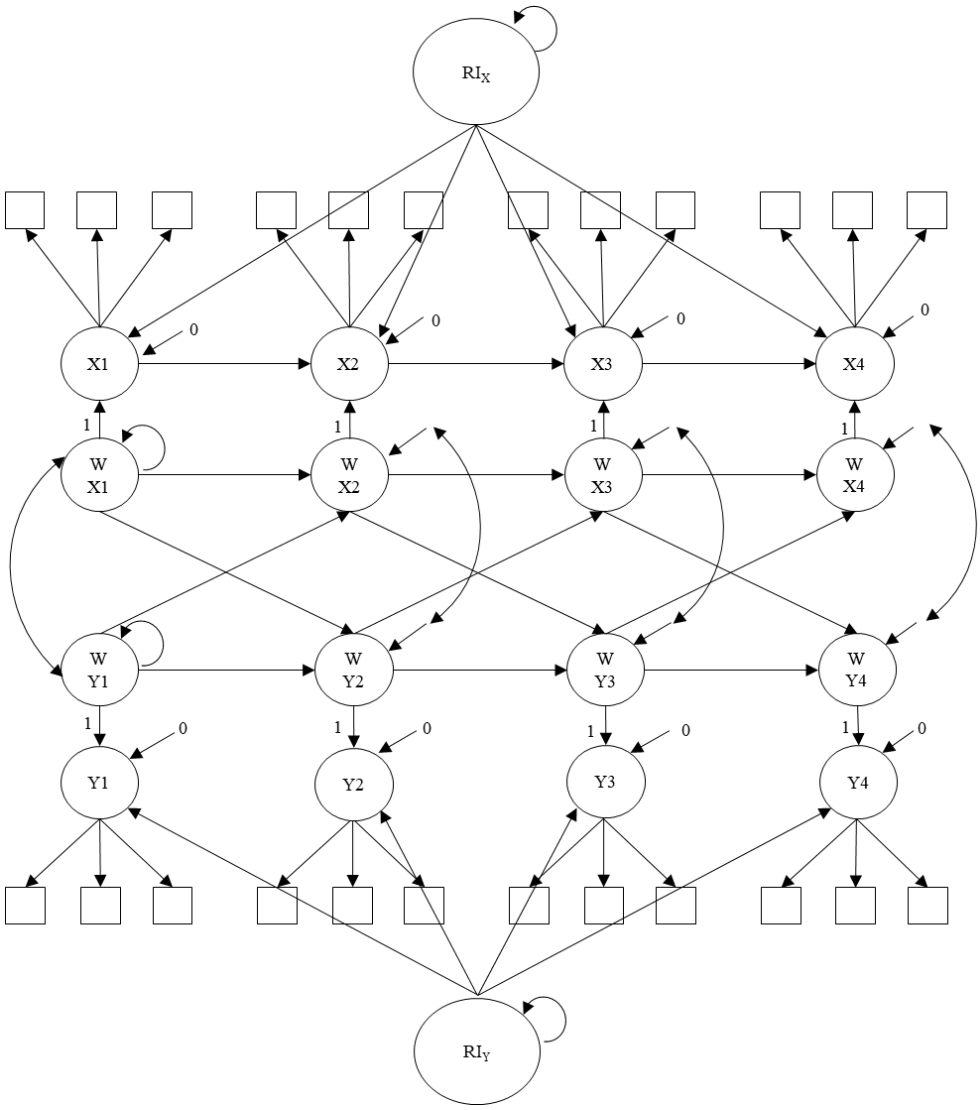
Multiple configurations of cross-lagged models with a random intercept have been proposed. The most prominent version is the random intercept cross-lagged panel model (RI-CLPM; Hamaker et al., 2015; see Figure 3, Panel B).[8] The RI-CLPM includes random intercept factors, which account for the common variance in the constructs over time, and these random intercept factors are correlated between constructs. Here, autoregressive and cross-lagged coefficients are estimated on the basis of the deviation from the typical level. A cross-lagged coefficient

[8] The RI-CLPM is equivalent to a bivariate Trait-State Occasion (TSO) model with autoregressive and cross-lagged effects (Cole, Martin, & Steiger, 2005; Eid, Holtmann, Santangelo, & Ebner-Priemer, 2017).

A) CLPM



B) RI-CLPM



C) LST-CLPM

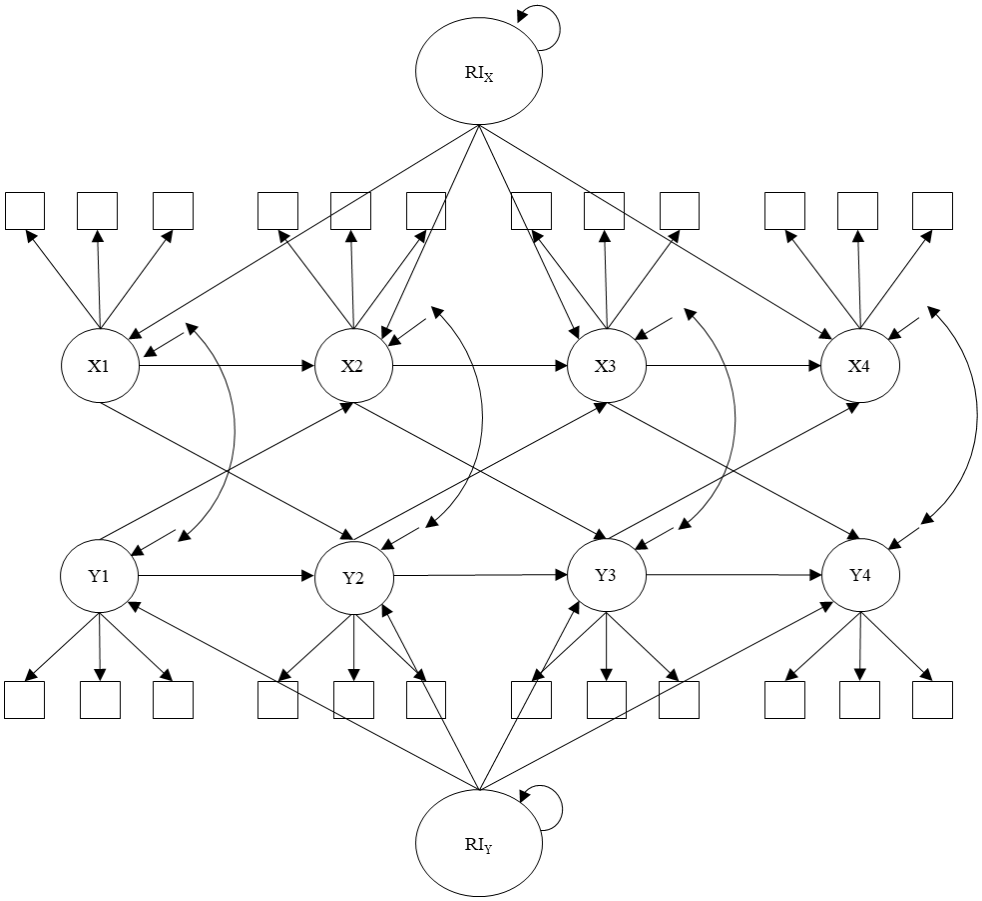


Figure 3. Simplified graphical representation of a four-wave CLPM, RI-CLPM, and LST-CLPM. The residual variances of the indicators are not displayed. In the RI-CLPM at T1, the within-person deviations (WX1 and WY1) technically have a variance; however, these are actually residual variances.

indicates whether individual differences in within-person deviations in one construct predict individual differences in within-person deviations in the second construct while controlling for previous individual differences in within-person deviations in the second construct (for an example, see Table 2). Thus, the RI-CLPM addresses the intraindividual level because within-person deviations are the entity of interest.

Another similar but not yet well-established model is the autoregressive cross-lagged model with unit effects (AR-CL model with units effects from Zyphur, Allison, et al., 2019), which puts autoregressive and cross-lagged paths on the latent (state) variables themselves while controlling for the random intercept factors. I will refer to this model as a latent state-

trait cross-lagged panel model (LST-CLPM; see Figure 3, Panel C) because it is a bivariate version of the autoregressive latent state-trait model (Steyer & Schmitt, 1994; Steyer, Mayer, Geiser, & Cole, 2015; for further explanation, see Study 2 of this dissertation). In line with the CLPM, this model addresses prospective relations on the basis of individual differences at a specific time point. However, in line with the RI-CLPM, the LST-CLPM controls for stable between-person differences in the outcome variable (for an example, see Table 2). More generally, the LST-CLPM explicitly points to the relation between latent state-trait models (Steyer et al., 1992) and cross-lagged models that include a random intercept. Whereas the random intercept factor represents the trait variance, within-person deviations represent the state residual variance from the latent state-trait model.

An overview of the characteristics of the CLPM, the RI-CLPM, and the LST-CLPM is presented in Table 2.[9] As indicated in Table 2, the models differ in (a) their conceptual idea, (b) whether they include a random intercept (i.e., trait variance), and (c) whether the cross-lagged (and autoregressive paths) are based on the latent state variance or the latent state residual variance. Subsequently, the interpretation of cross-lagged paths varies across the models. Another technical difference that is not displayed in the table is that the RI-CLPM and the LST-CLPM require at least three waves of data in order to be identified, whereas the CLPM requires only two waves of data. Along with this technical requirement, the number of waves can possibly affect the decomposition of the variance because the random intercept (i.e., trait variance) represents the common variance across all *available* waves. Yet, the more waves there are available, the more reliable the estimation of the random intercept (i.e., latent trait variance) will be (Rogosa, 1980; Singer & Willett, 2003). Correspondingly, a previous review of different cross-lagged panel models observed that models including random intercept factors had higher convergence rates when more waves of data were included (Orth et al., 2020). Overall, previous studies that compared these and other configurations of cross-lagged models based on real and simulation data observed that the CLPM showed the best convergence rates in comparison with other cross-lagged models (Orth et al., 2020; Usami, Todo, et al., 2019). Previous papers reviewing different cross-lagged models have made different recommendations to readers. Whereas some have implied that researchers should completely turn away from traditional cross-lagged panel models due to the confounding of within- and between-person variance

[9] Another similar model that includes random intercept factors is the STARTS model (Kenny & Zautra, 1995, 2001). The RI-CLPM can be considered a special case of the STARTS model (Hamaker et al., 2015). In the STARTS model, in addition to the RI-CLPM, the measurement error is modeled explicitly. However, the STARTS model is often difficult to estimate and requires a substantial number of waves.

(Hamaker et al., 2015; Usami, Murayama et al., 2019), others have argued that the CLPM should be used given the large interest in between-person effects (Orth et al., 2020).

Integrating methods and theoretical assumptions about teacher-student relationships and self-esteem. The different configurations of cross-lagged models presented above are based on different conceptual ideas about the underlying mechanisms, and they have different corresponding translations for modeling the complexity of longitudinal data. An important differentiation is whether researchers are interested in overall individual differences between persons or whether they are interested in within-person shifts in constructs as the driver of the underlying psychological process. Previous research on reciprocal relations between social relationships and self-esteem has almost exclusively relied on the traditional CLPM (Harris & Orth, 2019). One recent study that investigated the relation between family environment and self-esteem applied both the CLPM and the RI-CLPM (Krauss et al., 2019). The two models showed similar patterns of results, but more empirical problems (e.g., nonconvergence) occurred with the RI-CLPM. Theoretical models on the relations between social relationships and self-esteem were not specific enough to derive which of these two models should be addressed. For example, sociometer theory (Leary & Baumeister, 2000) suggested that social relationships gauge self-esteem; however, this can be based on two different perspectives. For example, in the relation between student-teacher relationships and self-esteem, the focus could be on either (a) individual differences in students' relationships with their teacher as predictors of individual differences in self-esteem or (b) students' deviations from the typical relationships with their teacher as the source of interest. The first focus is of particular relevance for those interested in understanding individual differences in student-teacher relationships and self-esteem. The second focus is particularly relevant from an intraindividual perspective, stressing the relevance of within-person deviations. From this perspective, the absolute level of student-teacher relationships and self-esteem is not of particular interest; rather, it is the individuals' deviations from their typical scores on these variables (independent of where on the continuum this deviation takes place). Comparing these could be particularly insightful in an extensive longitudinal design that can reliably estimate enduring trait differences.

Integrating cross-lagged models and latent state-trait models. As indicated above, different configurations of cross-lagged models are closely connected to latent state-trait models. For example, a random intercept factor reflects the trait factor that is present in the latent state-trait model. From this perspective, further research integrating these classes of models

could provide important insights into the specificities and differences between the different cross-lagged models. Moreover, because latent state-trait models are likely to be interwoven with state-trait measures (see Chapter 1.2), the joint consideration of cross-lagged models and the twofold state-trait operationalizations (state-trait measures and state-trait variance) could be used to further extend this integrative perspective. A particularly interesting context for such an undertaking represents the reciprocal relations between self-esteem and depressive symptoms. The vast amount of research on this relation has traditionally applied cross-lagged panel models (CLPMs) and has relied on trait measures of self-esteem and state measures of depressive symptoms. The results have indicated that trait self-esteem predicts state depressive symptoms (see Chapter 1.3.1). However, it is entirely unclear how state-trait variations in the measurement of self-esteem and depressive symptoms would impact the results on this longstanding research question, let alone the consequences of using cross-lagged models that have incorporate latent state-trait models (e.g., the LST-CLPM).

Table 2
Characteristics of Three Types of Cross-Lagged Panel Models

Model	Conceptual idea	Random intercept	Cross-lagged paths based on...	Exemplary interpretation
CLPM	Individual differences in X at T1 predict individual differences in Y at T2 (controlling for individual differences in Y at T1).	No	Latent state variance	Students with higher teacher support than others are predicted to have higher self-esteem at the next time point, controlling for prior self-esteem.
RI-CLPM	Individual differences in deviations from the typical level of X at T1 predict individual differences in deviations from the typical level of Y at T2 (controlling for individual differences in deviations in Y at T1).	Yes	Latent state residual variance	Students with deviations from the typical amount of teacher support that are higher than other students' deviations are predicted to have higher deviations from the typical level of self-esteem compared with other students' deviations while controlling for previous deviations from the typical level of self-esteem.
LST-CLPM	Individual differences in X at T1 predict individual differences in Y at T2 controlling for individual differences in the typical level of Y (and additionally controlling for individual differences in Y at T1).	Yes	Latent state variance	Students with higher teacher support than others are predicted to have higher self-esteem at the next time point, controlling for previous self-esteem and controlling for their typical level of self-esteem.

Note. CLPM = Cross-lagged panel model; RI-CLPM = Random intercept cross-lagged panel model; LST-CLPM = Latent state-trait cross-lagged panel model.

2 AIMS AND RESEARCH QUESTIONS

The present dissertation investigates emerging questions about global self-esteem by integrating substantive theoretical considerations and different methodological implementations. Global self-esteem is a central psychological construct as it describes individuals' overall feelings of worth, which are closely linked to psychological indicators of mental health (James, 1890/1963; Orth & Robins, 2014; Rosenberg, 1989). The study of self-esteem is of particular concern during adolescence and early adulthood because, in this phase, individuals have to face important challenges in life and are therefore more responsive to changes in self-esteem (Harter, 1998; Rosenberg, 1986). Specifically, the present dissertation addresses questions about the conceptualization, stability, and reciprocal relations of self-esteem. In doing so, the present dissertation builds on an integration of theory and methods. This approach is motivated by the aim to refine the translation of theoretical questions into their corresponding methodological operationalizations (Borsboom, 2006; Shadish et al., 2002). In fact, a substantive-methodological integration can create synergistic effects for a more granulated understanding of both theory and methods (Greenwald, 2012; Marsh & Hau, 2007). Therefore, this dissertation is grounded in two overarching objectives.

The first objective is to improve the understanding of self-esteem. For this purpose, I want to address three emerging areas of interest in research on self-esteem outlined in Chapter 1.4. First, I want to examine different conceptualizations of global self-concept as the apex of a multidimensional hierarchy (bottom-up vs. top-down) and set them in the relation to unidimensional global self-esteem. Second, I aim to study the state-trait implementation of self-esteem and its consequences for the relation to depressive symptoms. Third, I want to investigate within- and between-person effects of the reciprocal relations between student-teacher relationships and self-esteem over time.

The second objective is to improve the understanding of different methodological implementations stimulated by and empirically tested in research on self-esteem. The field of self-esteem research is a particularly fruitful environment for this undertaking because it is based on a large history of research, including a variety of theoretical and empirical work. I want to provide insights into the consequences of different operationalizations of higher order constructs, states and traits, as well as the consequences of using different cross-lagged models to analyze reciprocal relations. All of these methodological challenges are important far beyond

the scope of research on self-esteem, but more broadly, they are centrally relevant across for example educational, personality, and developmental psychology.

In order to integrate theory and methods in research on self-esteem, I draw on three empirical studies:

Study 1 (*Rethinking the Elusive Construct of Global Self-Concept: A Latent Composite Score as the Apex of the Shavelson Model*) merged theoretical and methodological perspectives on the conceptualization of global self-concept, which has been described as the apex of a multidimensional hierarchy of self-concept. We applied a reflective modeling approach using second-order factor models in which global self-concept affects lower order domain-specific self-concept, and we compared this approach with a more formative modeling approach that used a model-based latent composite score to capture the process by which global self-concept is formed on the basis of lower order self-concepts. In order to compare the consequences of applying each of the two approaches, the study evaluated internal criteria (i.e., correlations between lower order self-concepts, variances, stabilities) and external criteria (i.e., relations with other constructs) of the models in three independent samples with adolescents and young adults.

Study 2 (*How State and Trait Versions of Self-Esteem and Depressive Symptoms Affect Their Interplay: A Longitudinal Experimental Investigation*) addressed the question of the stability of self-esteem by using a two-fold operationalization of states and traits (measurement and modeling approach). Using an exploratory and a confirmatory experimental longitudinal study, university students were randomly assigned to state and trait measures of self-esteem (and depressive symptoms). State and trait measures were operationalized by using different time frames in the questionnaires ("In general…" vs. "During the last 2 weeks…"). Subsequently, the study examined the decomposition of state (residual) and trait variance in the different state/trait time frame conditions. In addition, the study examined how state-trait measures and the proportions of state-trait variance in self-esteem and depressive symptoms were related over time using cross-lagged panel models and latent state-trait cross-lagged panel models.

Study 3 (*Is Teacher Attachment Prospectively Related to Self-Esteem? A 10-Year Longitudinal Study of Mexican-Origin Youth*) investigated whether student-teacher relationships (i.e., students' attachment to their teachers) and self-esteem are reciprocally related over time. In order to study these reciprocal relations, the study applied traditional cross-lagged panel

models (i.e., focus on overall between-person differences) as well as random intercept cross-lagged panel models (i.e., focus on differences in within-person deviations). Besides capturing reciprocal relations during the years of schooling, the study extended the perspective to self-esteem development up to 4 years after the end of high school. The study used data from an extensive longitudinal data set of Mexican-origin youths living in California, including yearly assessments from age 11 to age 21.

In order to increase the transparency and robustness of the empirical studies, this dissertation adopted open science practices such as preregistration, replication, and open materials. Study 1 was exploratory in nature. Study 2 included both exploratory and confirmatory studies. Study 3 was mostly confirmatory. For the confirmatory studies, I preregistered the research questions and analytical procedures, which can be found on the Open Science Framework (links are provided in the studies). Studies 1 and 2 used multiple independent samples in order to increase the robustness of the results. More specifically, whereas in Study 1, multiple samples presented rather conceptual replications (e.g., by using different sets of self-concepts), Study 2 included a direct replication study that paralleled all the steps from the initial study. For all of the studies, I uploaded the syntax and output on the Open Science Framework (links are provided in the studies).

3 STUDY 1: RETHINKING THE ELUSIVE CONSTRUCT OF GLOBAL SELF-CONCEPT: A LATENT COMPOSITE SCORE AS THE APEX OF THE SHAVELSON MODEL

Braun, L., Rieger, S., Spengler, M., Göllner, R., Rose, N., Trautwein, U., & Nagengast, B. (2020). Rethinking the elusive construct of global self-concept: A latent composite score as the apex of the Shavelson model. *Manuscript submitted for publication.*

A preprint of this manuscript is available at the following link: https://psyarxiv.com/dbkw6

Abstract

The multidimensional, hierarchical model of self-concept by Shavelson et al. (1976) is a cornerstone of modern self-concept research. Given the comprehensive research interest in it, it is surprising that one core aspect of this model has yet to be clarified: What is the best way to operationalize the elusive construct of global self-concept as the apex of the hierarchy? Previous research implemented global self-concept with reflective modeling procedures (e.g., second-order factor models). Reflective models follow a top-down logic, which assumes that global self-concept affects lower order self-concepts. However, theoretical considerations often emphasize bottom-up processes, in which lower order self-concepts form a global self-concept. Yet, a bottom-up approach has not garnered much empirical interest, most likely because the requisite statistical models have not been available. The recently proposed model-based latent composite score can fill this gap. Therefore, we contrast top-down and bottom-up representations of global self-concept by comparing conventional second-order factors and model-based latent composite scores. Across three independent large-scale studies (Study 1: N = 8,068; Study 2: N = 3,876; Study 3: N = 2,095), our results indicate that composite scores have higher variances and a more plausible pattern of stabilities and correlations with external criteria (i.e., self-esteem, enjoyment of school, academic outcomes) than conventional second-order factors. In fact, the second-order factor model yielded smaller variances that boosted the correlations to theoretically and partially empirically implausible levels. We discuss the consequences of the two approaches and propose a latent composite score as the apex of the Shavelson model.

Keywords: multidimensional hierarchical self-concept, Shavelson model, global self-concept, second-order factor, model-based latent composite score

Rethinking the Elusive Construct of Global Self-Concept: A Latent Composite Score as the Apex of the Shavelson Model

Jasmin has a low overall opinion of herself. Likewise, she thinks she is not good at math or English and does not think she has a lot of friends. By contrast, Rafael thinks he has many friends and feels fairly positive about English; however, he thinks he is not good at math. Overall, Rafael usually thinks positive thoughts about himself. These are two examples from a large array of configurations of self-perceptions that can be found across students. These examples pose a crucial question that has drawn the interest of researchers for decades: How are students' more global and more specific self-perceptions related? For example, does Jasmin's low overall view of herself shape how she sees herself in different areas of life (e.g., in math)? Or do her perceptions of different domains (e.g., math or her social life) form her overall view of herself?

Just like Jasmin and Rafael, every student holds self-related perceptions in different areas of life and at different levels of specificity. These perceptions are typically referred to as *self-concepts*. The structure of self-concepts has been studied intensively across the last several decades of educational research. In this context, the *Shavelson model* (Marsh & Shavelson, 1985; Marsh, 1987; Shavelson et al., 1976) has become the most influential model (cited 5,680 times on Google Scholar in early 2020). The Shavelson model defines self-concept as a multidimensional (i.e., including self-concepts in domains such as math, English, or social contexts) and hierarchically ordered (i.e., more global and broader self-concepts are located at higher levels of the hierarchy) construct. There is comprehensive and convincing evidence for the multidimensionality of self-concept (Brunner et al., 2010; Marsh & Shavelson, 1985; Marsh, 1986b, 1987, 1990; Shavelson & Marsh, 1986), yet its hierarchical nature has puzzled researchers to this day. This is likely the case because an explicit, substantive definition of global self-concept was missing from the Shavelson et al. (1976) article. Another major reason for this might be that theoretical considerations about the hierarchy postulated by Shavelson et al. (1976) do not seem to match up with the subsequent methodological implementations, but such a match would be an important requirement for drawing valid conclusions (Greenwald, 2012; Marsh & Hau, 2007). As a consequence, it is still unclear whether global self-concept is best conceptualized as being formed by more specific self-concepts in a bottom-up way or whether global self-concept is best conceptualized as a high-level construct that affects more specific self-concepts in a top-down manner. A better understanding of global self-concept is important

(a) because of its centrality in this influential model, (b) for identifying and better understanding the emergence and long-term effects of global self-perceptions, and (c) for the potential to make decisions about starting points from which to successfully foster global self-concept.

Besides a lack of clarity regarding the definition of global self-concept, methodological restrictions have limited researchers to using top-down concepts of global self-concept. Thus, widely available higher order factor models, which follow a top-down logic, have dominated the analyses of the Shavelson model (Shavelson et al., 1976) and its extensions (Brunner et al., 2010; Marsh & Shavelson, 1985; Marsh, 1986b, 1987, 1990; Shavelson & Marsh, 1986). Until recently, alternative models that are more in line with the theoretical idea of bottom-up processes and can account for measurement error have not been available to applied researchers. Using newly developed latent variable models for latent composite scores (Rose, Wagner, Mayer, & Nagengast, 2019), we examined a bottom-up representation of global self-concept. We compared this new approach with conventional reflective second-order factor models (see, e.g., Marsh & Hocevar, 1985) using three independent large-scale studies (Study 1: N = 8,068; Study 2: N = 3,876; Study 3: N = 2,095) that contained different sets of self-concept measures. To evaluate the consequences and implications of the two approaches, we targeted different internal criteria (i.e., intercorrelations between self-concept measures, descriptives, and stabilities of global self-concept) as well as external criteria (i.e., correlations with different external criteria).

A Second-Order Factor Operationalization of Global Self-Concept: An Evolutionary Impasse?

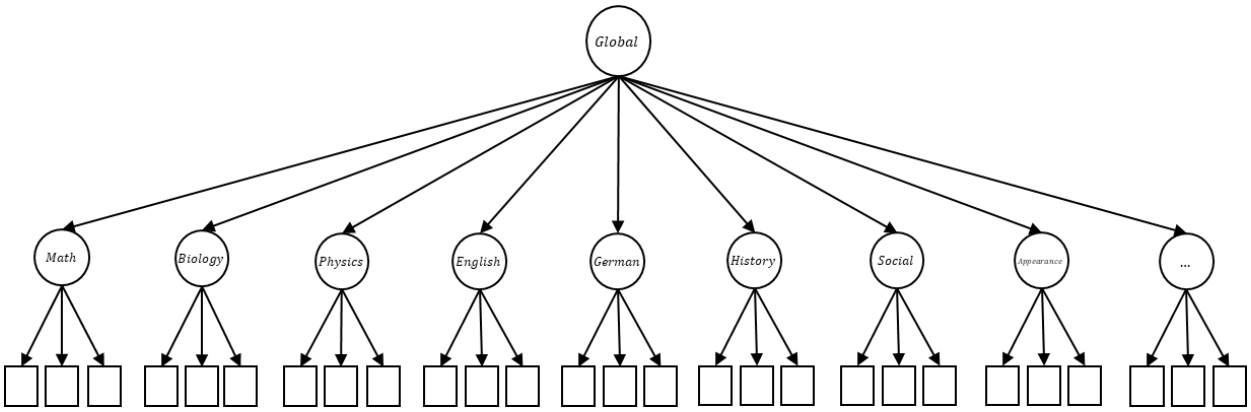
According to Marsh and Craven (1997) and extended by Brunner et al. (2010), there have been different evolutionary phases in research on the structure of self-concept, and they have made different assumptions about global self-concept. Until the 1970s, self-concept research appeared to be in a so-called *Dustbowl Phase* (Marsh & Craven, 1997) in which researchers used multiple self-concept items and summed them to obtain a global self-concept score. However, the composition and inclusion criteria for the self-concept items (e.g., Coopersmith, 1967; Fitts, 1965) lacked a solid theoretical and empirical foundation, and the methodological approaches reflected contemporary standards in that researchers did not account for measurement error or missing data.

To overcome the previous (theoretical) limitations of self-concept research, Shavelson et al. (1976) proposed a multidimensional (e.g., math, verbal, social) and hierarchically ordered

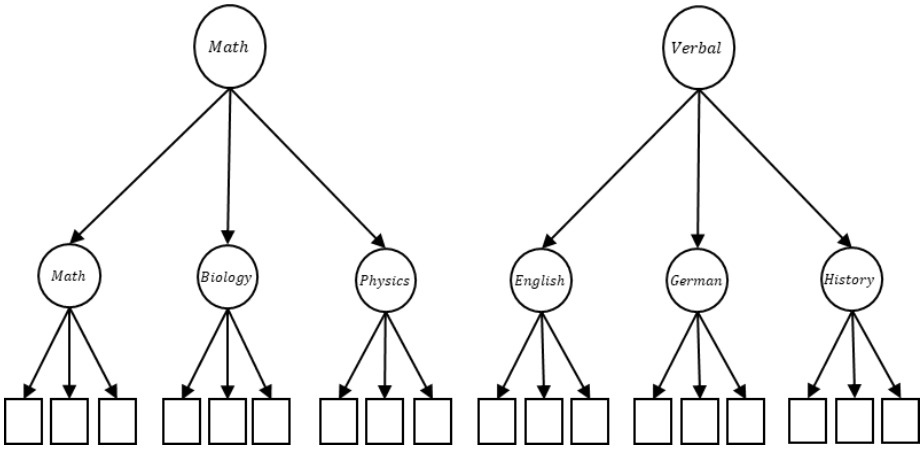
model of self-concept, which is commonly referred to as the *Shavelson model.* In this model, global (or general) self-concept is located at the apex of the hierarchy with academic and non-academic self-concept at the intermediate level and self-concepts in different subdomains (e.g., math or physical appearance) at the lowest level, which we will refer to as lower order self-concepts. However, Shavelson et al. (1976) made conflicting assumptions about the hierarchy. On the one hand, they posited that changes in higher order self-concepts require changes in lower order self-concepts, which implicitly suggests a bottom-up formation process within the hierarchy of self-concept. On the other hand, they compared the hierarchy of self-concept with the g-factor approach from the hierarchical structure of intellectual abilities (see, e.g., Spearman's g-factor; Soares & Soares, 1977; Vernon, 1950). The latter consideration was implemented via confirmatory factor analysis, which assumes a reflective latent variable model that, if interpreted in a realistic manner, explicitly implies a top-down process (Borsboom, Mellenbergh, & van Heerden, 2003). Even if theoretically emphasizing a bottom-up process within the hierarchy of self-concept, Marsh and Shavelson (1985) jumped on the confirmatory factor analysis bandwagon to implement global self-concept. This methodological operationalization was a path-breaking decision as it has dominated research on self-concept to this day.

However, the first studies to use second-order confirmatory factor analysis (see Figure 1, Part a) had already unfolded difficulties because second-order factor models did not fit the data well, and global self-concept (represented as the second-order factor) tended to have a small variance, which resulted from low correlations between math and verbal academic self-concepts (Brunner, Lüdtke, & Trautwein, 2008; Marsh & Hocevar, 1985; Marsh & Shavelson, 1985; Marsh, 1986b, 1987, 1990; Marsh & Hau, 2004; Shavelson & Marsh, 1986). These low correlations even decreased with age (Marsh, 1989; Marsh, Craven, & Debus, 1991; Marsh & Ayotte, 2003). Notwithstanding the empirical problems of the second-order factor model (i.e., a small variance based on low correlations between academic self-concepts), its results have been widely interpreted. For example, a second-order factor for global self-concept and unidimensional global self-esteem were highly correlated, which led researchers to conclude that they probably represented the same construct (Marsh & Hattie, 1996). However, this conclusion was made without acknowledging potential problematic parameter estimates due to the small variances of the second-order factor.

(a)



(b)



(c)

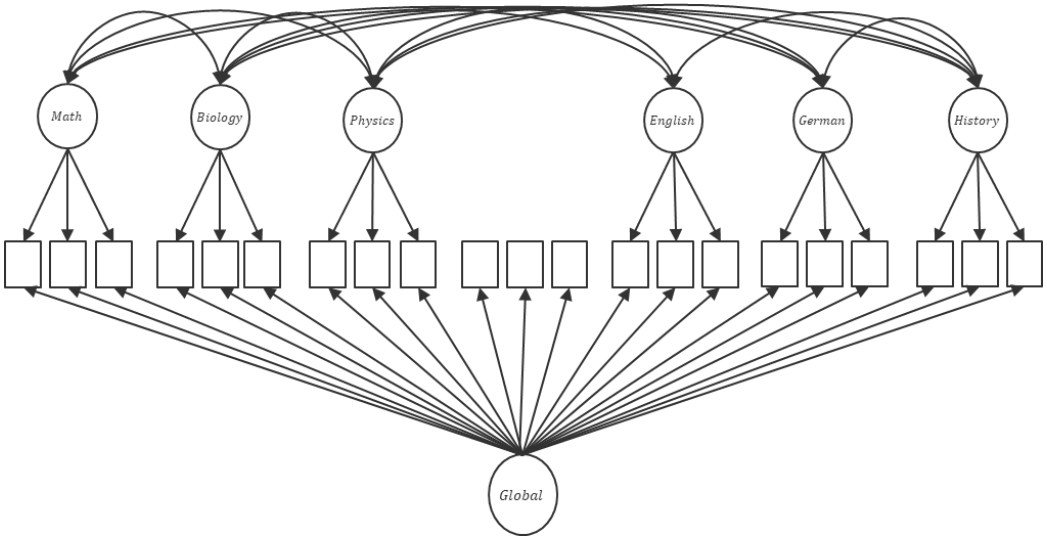


Figure 1. Three representation of the structure of self-concept: (a) second-order factor based on the Shavelson model; (b) Marsh/Shavelson model; (c) Nested Marsh/Shavelson model.

At the same time, Marsh and Shavelson (1985) called the existence of global self-concept into question more generally. Accounting for the low correlations, they proposed that academic self-concept be separated into two uncorrelated factors representing math and verbal academic self-concept (see Figure 1, Part b), and they consequently abandoned global self-concept. This model is known as the *Marsh/Shavelson model* (Marsh, 1990). Driven by further methodological developments in confirmatory factor analysis (Eid, Lischetzke, Nussbeck, & Trierweiler, 2003), Brunner and colleagues postulated the *Nested Marsh/Shavelson model* for academic self-concepts (see Figure 1, Part c), which separates general academic self-concept from domain-specific (method) factors using a correlated trait correlated method minus one (CT-C[M-1]) approach (Brunner et al., 2008; Brunner, Keller, Hornung, Reichert, & Martin, 2009; Brunner et al., 2010; Eid et al., 2003). The (more complex) Nested Marsh/Shavelson model fit the self-concept data better than previous versions (Brunner et al., 2008; Brunner et al., 2009; Brunner et al., 2010). However, in line with the second-order factor model, this approach followed reflective modeling approaches, which imply a top-down process for global (academic) self-concept.

In sum, the Shavelson model stimulated important developments on research on the structure of self-concept. Theoretically, Shavelson et al. (1976) proposed a bottom-up approach within the hierarchy of self-concept, yet at the same time, they made a conflicting argument when drawing on the g-factor analogy, which underlies a top-down approach. Along with elaborated methodological developments such as (second-order) confirmatory factor analysis, this led to a dominance of reflective models when studying the Shavelson model. These models imply a top-down logic rather than the theoretically proposed bottom-up idea. Therefore, in the present research, we asked whether previous developments also framed as "evolutionary step[s]" (Brunner et al., 2010, p. 976) have run into an evolutionary impasse instead? One reason for this impasse might be that the theoretical ontology underlying these statistical models was not sufficiently merged with theoretical considerations on the hierarchy of self-concept (Greenwald, 2012; Marsh & Hau, 2007). Hence, considerations of whether a top-down or a bottom-up approach is appropriate for representing the structure of self-concept call for a deeper understanding of their empirical implementations (see, e.g., Bollen & Lennox, 1991; Bollen & Diamantopoulos, 2017; Edwards & Bagozzi, 2000; MacKenzie, Podsakoff, & Jarvis, 2005).

Assumptions of Reflective and Formative Modeling Approaches

When choosing the most appropriate statistical (measurement) models, it is of central importance to achieve a good match with the assumed theoretical ontology of a construct (Borsboom et al., 2003, 2004). In the following, we therefore reviewed the statistical assumptions of the two modeling approaches representing a top-down versus a bottom-up implementation of the hierarchy of self-concept. It is important to note that (a) we focused on the first-order level (i.e., the relation between lower order self-concepts and global self-concept) and not on the lowest level (i.e., the relation between item indicators and first-order factors), and (b) we addressed the directional link *within* the self-concept hierarchy in order to model global self-concept and not the temporal predictions between different hierarchically ordered self-concepts over time (for this approach, see Harris, Wetzel, Robins, Donnellan, & Trzesniewski, 2018; Marsh & O'Mara, 2008; Rentzsch & Schröder-Abé, 2018; Trautwein, Lüdtke, Köller, & Baumert, 2006).

Reflective models are the most commonly used measurement models in psychological research (Borsboom et al., 2003). In reflective models, higher order constructs are causes of indicators (in our case, lower order self-concepts), often referred to as effect indicators (Blalock, 1964). These models presume a causal relation from a higher order (inferred) construct to its indicator in such a way that a change in the (higher order) construct determines a change in the indicators (Edwards & Bagozzi, 2000). In general, the higher order construct is modeled to represent a unidimensional construct, which calls for moderate to high positive intercorrelations between its indicators and assumes that the indicators should be interchangeable (Bollen & Lennox, 1991). Analyses of reflective models representing global self-concept have run into difficulties, most likely because of the low correlations between the lower order self-concepts that have been in conflict with the assumptions of reflective models (Bollen & Lennox, 1991). This makes a reflective model questionable because—by definition—it can explain only a small proportion of common variance in the lower order constructs. Indeed, Marsh (1987) already noted that previous results on the hierarchy of self-concepts

> “ […] have more negative consequences for a top-down model that makes the stronger theoretical assumption that the hierarchical general self "causes" the lower order facets. If a hierarchical general self is posited to cause a lower order factor but the two are nearly uncorrelated, then the postulated causal relation is not supported.” (Marsh, 1987, p.34)

As a consequence, on the basis of a systematic and rigorous application of state-of-the-art methodology, previous self-concept research has concluded that the top-down order must

be “more complex” (Marsh & Shavelson, 1985, p. 121). However, what was missing back then were alternative statistical models that would have allowed for a sophisticated formative bottom-up modeling of global self-concept.

In fact, the causal flow of formative modeling approaches is the opposite of reflective models. Here, indicators cause higher order constructs (Bollen & Lennox, 1991; Edwards & Bagozzi, 2000) and are therefore called cause, causal, formative, or composite indicators (Blalock, 1964; Bollen & Lennox, 1991). Accordingly, the higher order construct is formed by the indicators. This requires that all indicators that form the construct are included in the model (Bollen & Lennox, 1991). In contrast to reflective models, removing an indicator has an impact on the nature of the construct. Furthermore, formative models do not require high intercorrelations between the indicators. The variance of the higher order construct is based on the conglomerate of the variances of the indicators.

So far, there has been little research on a formative perspective on the hierarchy of self-concept. The few existing studies go back to the time before the Shavelson model was introduced (e.g., Coopersmith, 1967; Hishiki, 1969; Ludwig & Maehr, 1967; Sears et al., 1972; Soares & Soares, 1969; Zirkel, 1971). At that point, global and domain-specific self-concept were modeled using the simple sum of a series of self-concept items that addressed self-concept in different areas of life. However, these studies lacked a clear theoretical and methodological foundation.

Historically, formative approaches have been criticized because of the difficulty of identifying the models (i.e., because they require a predicted variable) and because they ignore measurement error (Bollen & Diamantopoulos, 2017), which has dramatic consequences for results in psychological research (e.g., Cole & Preacher, 2014). Correspondingly, this led to a boost in reflective (measurement) models across psychology and beyond. These models were relatively easy to implement and accounted for measurement error in the observed variables (Borsboom, 2008; Rhemtulla, van Bork, & Borsboom, 2019). However, more recent methodological considerations have emphasized that model choice should not be driven by the ability to handle measurement error but by the theoretical ontology of a psychological construct (Rhemtulla, van Bork, & Borsboom, 2015, 2019). The misplaced application of reflective (measurement) models can result in the invalidity of the construct, model misfit, and biased structural estimates (Rhemtulla et al., 2019). Moreover, Rose et al. (2019) developed a formative approach for obtaining latent composite scores that are free from measurement error and

can be implemented easily[10]. Hence, the central idea guiding this research was to examine both a top-down approach by using reflective models and a bottom-up approach by using formative models to model global self-concept.

The Present Study

What is the best way to operationalize the elusive construct of global self-concept as the apex of the self-concept hierarchy? Is it formed by different lower order self-concepts, or is it the cause of different lower order self-concepts? In the present study, we addressed this question by comparing the consequences of a top-down versus a bottom-up representation of global self-concept as the apex of the Shavelson model (Shavelson et al., 1976). In empirical studies, the Shavelson model was nearly exclusively considered a top-down model because it was based on reflective modeling procedures. However, this was most likely a result of methodological rather than theoretical considerations. In addition, previous results on the Shavelson model when reflective models (i.e., second-order factor models) were used revealed difficulties because the lower order self-concept factors were barely correlated, which contradicts the crucial assumption of a common underlying cause in reflective models. In the present study, we provide insights into the consequences of a bottom-up representation of the Shavelson model using a formative modeling procedure. To do this, we analyzed three independent longitudinal large-scale studies that included different sets of self-concepts and compared the consequences of the reflective and formative representations of the Shavelson model by drawing on the internal (i.e., properties within the model) and external (i.e., relations to other constructs) criteria of the models.

Internal criteria. For the internal criteria, we addressed three overarching research questions. First, we analyzed the intercorrelations between the lower order self-concepts (Research Question 1). This pattern of intercorrelations is particularly interesting because it lays the foundation for further modeling decisions (e.g., Bollen & Lennox, 1991). We expected to replicate the findings of previous studies that showed that (at least some of the) first-order self-

[10] There is some ambiguity in the literature about the term *formative* (Bollen & Bauldry, 2011). Whereas some researchers use the term formative indicators in order to refer to composite indicators, others use it in reference to causal indicators that include a disturbance term. In the present study, we refer to composites as part of the conceptual idea of formative models and, therefore, we do not further distinguish the terms composite and formative indicators (respectively composite and formative models). In addition, it is important to note that we focus on the formative *second*-order level and not on the formative *first*-order level (see e.g., Diamantopoulos, Riefler, & Roth, 2008).

concepts are barely correlated (Marsh & Shavelson, 1985). Second, we investigated whether there were meaningful differences in the variance of global self-concept when it was modeled by the reflective versus the formative approach (Research Question 2). Third, we analyzed whether the temporal stabilities of global self-concept differed between the reflective and formative approaches (Research Question 3). Shavelson et al. (1976) considered the stability of different hierarchical elements to be one important feature of his model and assumed that global self-concept should be very stable over time.

External criteria. Along with these internal criteria, we analyzed the relation between global self-concept and external criteria in order to gain a broader understanding of the consequences of the two approaches. In the convergent validity framework (Campbell & Fiske, 1959), we asked how global self-concept (operationalized by the two approaches) is related to theoretically similar constructs (Research Question 4). We chose different target constructs: Primarily, we expected that global self-concept would be positively related to global self-esteem (Marsh & Shavelson, 1985), which was directly measured (Rosenberg, 1965) with a standard questionnaire (Studies 1 and 2). Self-esteem typically refers to a unidimensional construct that is defined as the global, subjective evaluation of one's own worth (Donnellan, Trzesniewski, & Robins, 2011). In previous studies, global measures of self-esteem and an inferred second-order factor based on different self-concept measures were highly correlated (Marsh & Hattie, 1996). Yet it is unclear what caused these high correlations given the small variances of the second-order factor models. Second, we analyzed the relation between global self-concept and students' enjoyment of school (Study 1), which is supposed to be an important outcome of students' adjustment and is a construct that was previously used to evaluate the construct validity of self-concept measures (Marsh & O'Neill, 1984). Finally, we analyzed the relations between global academic self-concept and global academic outcomes. Previous research had shown that self-concepts are reciprocally related to academic outcomes in the same domain (Marsh & O'Neill, 1984; Marsh & Yeung, 1998). Following the matching specificity principle (Swann, Chang-Schneider, & Larsen McClarty, 2007), we identified outcomes that were on a comparable specificity level as global self-concept. Therefore, we chose to analyze very global academic outcomes, such as an average achievement score, grade point average, and the transition after secondary school.

In order to empirically model the reflective approach, we used traditional second-order factor models. For the formative approach, we used a newly developed modeling procedure, namely, the model-based-latent composite score (Rose et al., 2019). The main advantages of

this approach compared with classical formative models are that (a) it can model composites of first-order factors, which are free from measurement error, (b) missing data can be considered by means of full information maximum likelihood estimation, (c) it allows the implementation of a formative composite variable independent of outcome-dependent weights, and (d) it can be implemented easily in conventional statistical software for structural equation models (e.g., Mplus or R; see Rose et al., 2019).

Method

The data sets used in the present study came from three multiconstruct large-scale studies. The data were collected in adherence with ethical principles in the treatment of sensitive personal data and were approved by the responsible school authorities (Study 1) and by the ministries of education and cultural affairs of the German states of Saxony (Study 2) and Baden-Württemberg (Studies 2 and 3).

Samples and Instruments

Study 1: BIJU. We analyzed data from the German multicohort longitudinal study "Learning Processes, Educational Careers and Psychosocial Development in Adolescence and Young Adulthood" (BIJU; Baumert et al., 1996). We used data from students in the first cohort at the beginning and end of Grade 7 (T1: N = 5,948; T2: N = 6,263) from academic and nonacademic schools in four German federal states (North Rhine-Westphalia, Saxony-Anhalt, Mecklenburg Western Pomerania, and Berlin [Berlin participated only at T2]). The total sample included N = 8,068 students (53% female), who participated at a minimum of one time point. The study incorporated the assessment of eight self-concepts, including five academic (Jerusalem, 1984; Jopt, 1978) and three nonacademic self-concepts (Fend & Prieser, 1986). In Table 1, the self-concept measures and their internal consistencies are displayed in detail. In sum, the self-concept measures consisted of three to four items and showed good reliabilities (αs > .70). Self-esteem was measured with a four-item German version (Jerusalem, 1984; Trautwein, 2003) of the Rosenberg Self-Esteem Scale (Rosenberg, 1965). Previous analyses showed that the short and the long versions of the Rosenberg scale were strongly correlated (Trautwein, 2003). The self-esteem scale showed good reliabilities (αs > .75). Furthermore, enjoyment of school was measured with three items (αs > .65) that assessed the overall pleasure and enjoyment of school (e.g., "I like going to school"). All items were rated on a 4-point Likert scale ranging from 1 (*totally disagree*) to 4 (*totally agree*).

Study 2: TRAIN. We analyzed data from the German longitudinal school achievement study "Tradition and Innovation" (TRAIN; Jonkmann, Rose, & Trautwein, 2013). In total, we considered data from *N* = 3,876 students (45% female) in Grade 5 (T1 = 2,101) and Grade 8 (T2 = 2,382) from 136 classes in 99 nonacademic schools in two federal German states (Baden-Württemberg, Saxony). In the TRAIN study, five self-concepts were assessed at both time points, including three academic self-concepts (math, German, English) and two nonacademic self-concepts (social, assertiveness). Table 1 presents an overview of the measures and their internal consistencies. In sum, the measures contained three to four items and showed good reliabilities (αs > .66). All self-concept measures were rated on a 4-point Likert scale ranging from 1 (*totally disagree*) to 4 (*totally agree*). Self-esteem was measured with the four-item self-esteem scale from the KINDL-R scale (Ravens-Sieberer et al., 2001), which assessed students' self-esteem (e.g., "In the last week, I was proud of myself) using a 5-point Likert scale ranging from 1 (*never*) to 5 (*always*) and showed good reliabilities (αs > .71).

Study 3: TOSCA-10. We employed data from the German longitudinal study "Transformation of the Secondary School System and Academic Careers: Grade 10" (TOSCA-10; Trautwein, Nagy, & Maaz, 2011). The study was designed to investigate the transition of non-academic intermediate 10th grade students to further education. At the first time point in Grade 10, data from *N* = 2,095 students (51% female) were assessed. At the second time point, *N* = 473 students could be assessed again, which took place 6 years later in order to assess students' educational trajectories. The study included five academic self-concepts (math, verbal, English, economic, technical) with four to eight items each that were rated on a 4-point Likert scale ranging from 1 (*totally disagree*) to 4 (*totally agree*) and showed good reliabilities (αs > .76; for details, see Table 1). Additionally, grade point average and an average standardized achievement score (German, English, math, biology, economics, technology) in Grade 10 and the educational trajectories 7 years later were assessed. Thereof, we used an indicator of whether students chose the academic (*N* = 188) or vocational track (*N* = 285) after Grade 10.

Statistical Analyses

Second-order factor model (SOF). For the reflective approach, we used traditional second-order factor models (see the lower left part of Figure 2). First, $\eta_1 \ldots \eta_Q$ first-order factors were modeled for all self-concept measures. Second, a second-order factor ξ was modeled in order to explain the covariation between the first-order factors. All first-order factors as well

Table 1
Overview of the Self-Concept Measures in the Three Studies

	Study 1: BIJU			Study 2: TRAIN			Study 3: TOSCA-10		
Self-concept	Items	α (T1, T2)	Example	Items	α (T1, T2)	Example	Items	α (T1, T2)	Example
Academic									
Math	4	.85, .90		4	.78, .86		4	.89, .91	"I am good at mathematics."
German/Verbal	4	.77, .87	"I don´t like mathematics/ German/English/ Biology/Physics very much."	4	.64, .66	"I am good at mathematics /German /English."	4	.76, .80	"I can express myself verbally well."
English	4	.83, .88		4	.68, .84		4	.92, .92	"I am good at English."
Biology	4	.87, .88							
Physics	4	.89, .89							
Economic							4	.84, .89	"I understand economic affairs well."
Technical							8	.94, .93	"I find it easy to solve technical problems."
Nonacademic									
Social	4	.69, .69	"I am quite respected among my classmates."	4	.70, .83	"I am quite respected among my classmates."			
Appearance	4	.47, .39	"I am quite satisfied with my physical development"						
Assertiveness	3	.57, .62	"I don't think I can assert myself as well as other people."	3	.66, .76	"I don't think I can assert myself as well as other people."			

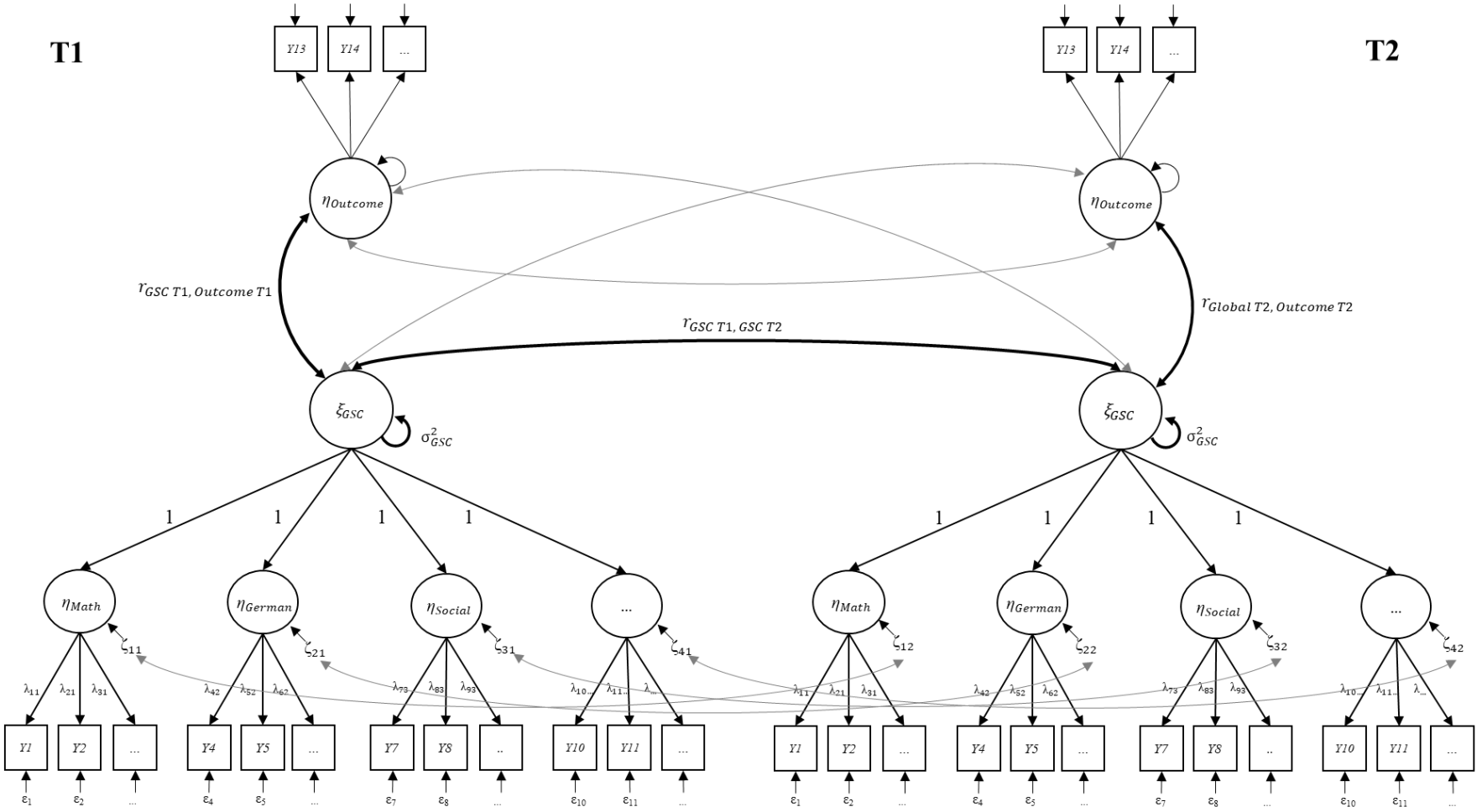


Figure 2. Simplified representation of the present longitudinal second order factor model. GSC = Global self-concept. Correlated uniquenesses are not displayed. Bold lines represent research questions.

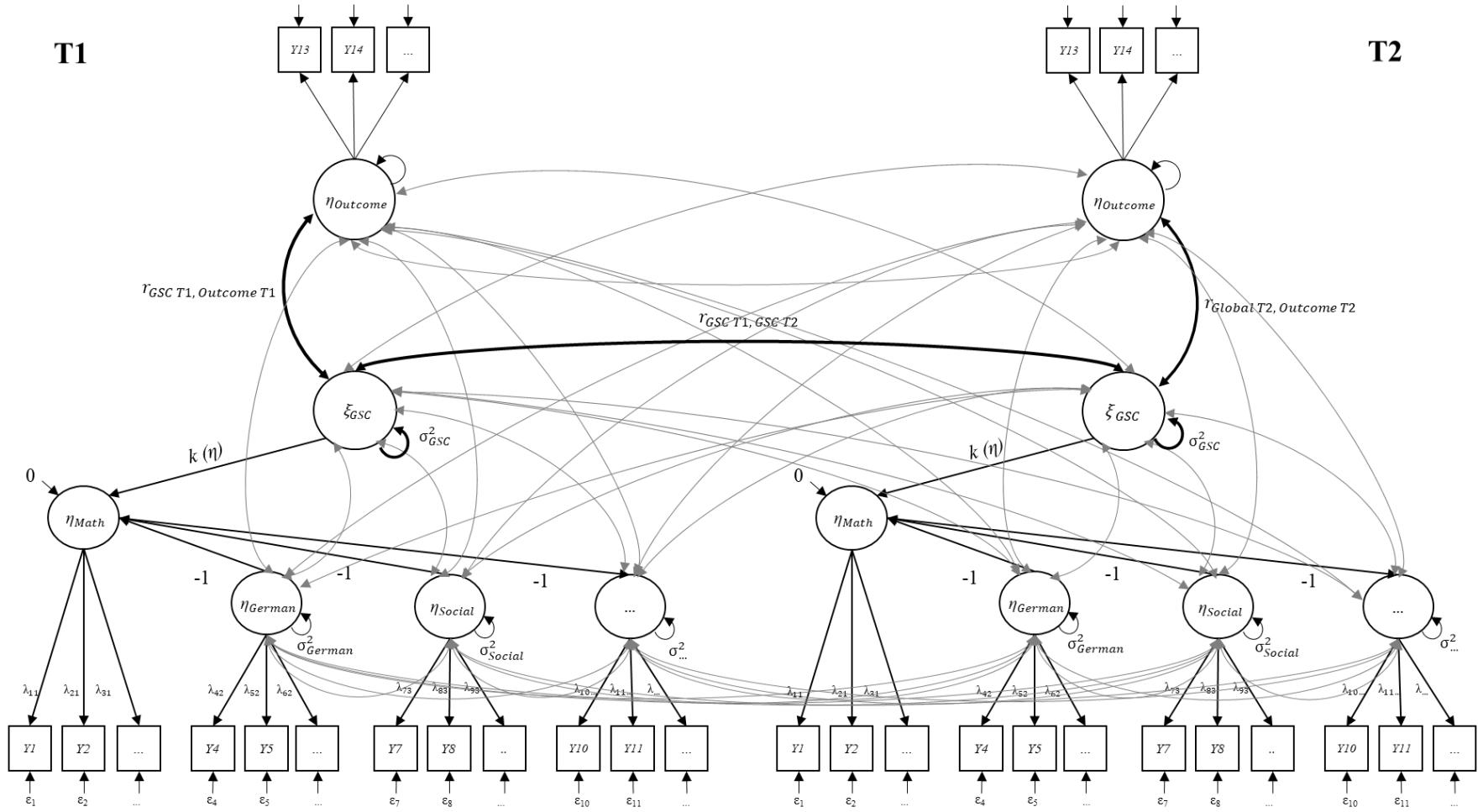


Figure 3. Simplified representation of the present longitudinal model-based composite score model. GSC = Global self-concept. Correlated uniquenesses are not displayed. Bold lines represent research questions.

as the second-order factor were identified by a "nonarbitrary" model identification (Little, Slegers, & Card, 2006). This identification approach applies a weighted combination of all indicators to ensure an optimal balance across indicators. Thus, through an average intercept of zero and an average factor loading of 1, the latent variable obtains an optimally weighted metric across all indicators

Model-based latent composite score (LCS). For the formative approach, we used model-based latent composite scores (see the lower left part of Figure 3). The LCS is a model that was recently developed by Rose et al. (2019). It allows composite scores to be estimated on the basis of first-order factors that account for measurement error. We chose to model an *average* composite score (for a sum score, see Rose et al., 2019). The model is similar to a latent change score model, except that an average instead of a difference is modeled. First, $\eta_{1...}$ η_{Q} first-order factors are modeled for all components of the composite score (all self-concept measures). In line with the second-order factor model, they were identified by a nonarbitrary model identification. In the next step, one of the first-order factors was arbitrarily chosen as the pseudo-indicator variable $\eta_{PIM,}$ which was modeled as the indicator of the composite score with a factor loading of the number of first-order factors and a fixed residual of 0. The paths of the other $\eta_{1...}$ η_{Q-1} first-order factors on the pseudo-indicator variable η_{PIM} were fixed to -1. By doing so, all components were equally weighted in the composite score (for a weighted composite score, see Rose et al., 2019). The latent LCS incorporates full information maximum likelihood estimation in order to adjust for item nonresponses and missing data in manifest variables.

Analytical procedure. In order to analyze the consequences of using a reflective versus a formative approach to assess the hierarchy of self-concept, we drew on the internal and external criteria from the modeling approaches used in three different longitudinal studies. In a first step, we analyzed the latent correlations between the lower order self-concepts. Second, we modeled global self-concept separately by applying the SOF and the LCS by using a stepwise inclusion of the self-concept measures and by computing its descriptives, its stabilities, and its relations to external criteria (see Figures 2 and 3). All models included both time points with strong measurement invariance imposed across time in the measurement models (Meredith, 1993). In addition, in the second-order factor model, we assumed strong measurement invariance across time in the structural model (i.e., loadings and intercepts of the first-order factors).

We addressed three aspects of second-order factor and composite score models, which are displayed in bold lines in Figures 2 and 3. First, we investigated the variances of global self-concept ($\hat{\sigma}^2_{GSC}$). Second, we investigated the temporal stability, which is represented by the correlation coefficients between the measurement time points across 1 (Study 1), 4 (Study 2), and 7 (Study 3) years ($r_{GSC\ T1,\ GSC\ T2}$). Third, we analyzed external relations by computing the correlations of global self-concept with global self-esteem and enjoyment of school (Studies 1 and 2) and global academic outcomes (Study 3), that is, average achievement, grade point average, and the transition after Grade 10 ($r_{GSC,\ Outcome}$). In addition to the correlations, we present covariances because the constructs have the same measurement units, and the correlations are highly influenced by the variances of the constructs.

All models were estimated in Mplus 8.2 (Muthén & Muthén, 1998-2017). All data analysis scripts are available for review on the Open Science Framework at the following address: https://osf.io/63nk8/?view_only=0e61f97dda0b48d4bd4d10eb463a9627 (the project will be publicly available after the review process has been completed). Due to specific item variance, we allowed for correlated uniquenesses (a) of the same items over time, (b) across items with the same wording, and (c) for reversed item formulations (Cole, Ciesla, & Steiger, 2007). Due to missing values on single items and over time, we used full-information maximum likelihood estimation (FIML). In order to rule out the possibility that different sets of missing data were responsible for differences between particular models, we included the same set of variables in all models. Thus, the FIML estimation was held constant across all models, and the models differed only in the specification of the structural model (i.e., which self-concepts were included in the second-order factor/composite score) but not concerning the measurement models and the covariances with the outcome variables. We took the nested data structure into consideration (i.e., students nested within classes) by using cluster-robust standard errors (McNeish, Stapleton, & Silverman, 2017; Muthen & Satorra, 1995). Across all three studies, longitudinal measurement models including all correlated first-order self-concepts fit the data well (CFI > .935, TLI > .922, RMSEA > .036; SRMR > .048; see the Appendix). The specification of the model-based LCS did not affect the model fit, as there are no additional implications regarding the means or the covariance structure on other variables in the model. Hence, the model with the LCS always had the same degrees of freedom and the same model fit as the model with correlated first-order factors (Rose et al., 2019). The longitudinal second-order factor models showed mostly acceptable model fits across all three studies (CFI > .917, TLI > .903, RMSEA

> .040, SRMR > .122; see the Appendix). The fits of all the models, including the models with external criteria, are presented in the Appendix.

Results

Internal Criteria

Before investigating global self-concept, we began by analyzing the latent correlations between the lower order self-concepts in the three studies (Research Question 1). Tables 2, 3, and 4 depict the intercorrelations between the lower order self-concepts in the three studies at two time points each. The results indicated that, across all three studies, most of the correlations were low (r < .30) and positive. However, correlations among conceptually distinct domains (e.g., math and verbal) were close to zero or even negative. In addition, some correlations among conceptually similar domains (e.g., English with verbal; social with appearance) were medium to high (r > .30). Correlations between the same lower order self-concept over time (i.e., rank-order stabilities) were medium to high (Study 1: r = .39 to .59; Study 2: r = .34 to .50; Study 3: r = .47 to .79).

In the next step, we specified second-order factor (SOF) and latent composite score (LCS) models that represented global self-concept. To be more specific, we examined different numbers and combinations of lower order self-concepts in order to gain an understanding of the dependencies and sensitivities in the two modeling approaches. First, we were interested in the variances (i.e., interindividual differences) of the respective global self-concepts (Research Question 2). As indicated in Table 5, across all three samples, all sets of self-concepts, and all time points, the composite scores exhibited larger variances ($\widehat{\sigma}^2_{LCS}$ = 0.10 to 0.33) than the second-order factors, which even yielded negative estimates ($\widehat{\sigma}^2_{SOF}$ = -0.12 to 0.26). More specifically, in Studies 1 and 2, the differences in the variances of global self-concept were larger when academic self-concepts were included ($\widehat{\sigma}^2_{LCS}$= 0.16 to 0.30; $\hat{\sigma}^2_{SOF}$ = 0.06 to 0.13). In Study 3, which included only academic self-concepts, the variance of a second-order factor based on only two self-concepts yielded a negative variance ($\widehat{\sigma}^2_{SOF}$ = -0.12), and the second-order factor model based on three self-concepts did not converge. Not surprisingly, the means of the second-order factor and the composite score did not differ substantially because the first-order factors were identified by the same identification method.

Table 2
Latent Intercorrelations of the First-Order Self-Concepts in the BIJU Study

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Math T1	-														
2. German T1	.21	-													
3. English T1	.16	.24	-												
4. Biology T1	.15	.31	.17	-											
5. Physics T1	.41	.22	.18	.45	-										
6. Social T1	.16	.25	.27	.20	.13	-									
7. Appearance T1	.29	.19	.24	.21	.32	.67	-								
8. Assertiveness T1	.20	.23	.22	.18	.21	.63	.55	-							
9. Math T2	.51	.13	.12	.14	.32	.13	.25	.16	-						
10. German T2	.09	.49	.15	.24	.17	.20	.14	.15	.23	-					
11. English T2	.14	.22	.58	.14	.18	.21	.19	.11	.17	.28	-				
12. Biology T2	.17	.17	.11	.39	.25	.14	.15	.09	.27	.29	.15	-			
13. Physics T2	.31	.12	.14	.19	.44	.10	.24	.14	.40	.21	.20	.41	-		
14. Social T2	.08	.13	.16	.12	.06	.53	.30	.31	.19	.23	.22	.21	.15	-	
15. Appearance T2	.19	.12	.18	.10	.21	.38	.59	.30	.28	.18	.25	.20	.28	.73	-
16. Assertiveness T2	.14	.15	.14	.12	.14	.43	.31	.52	.24	.23	.22	.24	.22	.77	.72

Note. N = 8,068.

Table 3
Latent Intercorrelations of the First-Order Self-Concepts in the TRAIN Study

	1	2	3	4	5	6	7	8	9
1. Math T1	-								
2. German T1	.29	-							
3. English T1	.18	.45	-						
4. Social T1	.27	.31	.33	-					
5. Assertiveness T1	.34	.39	.32	.65	-				
6. Math T2	.42	-.00	.01	-.01	.05	-			
7. German T2	.18	.50	.29	.15	.14	.22	-		
8. English T2	.08	.27	.36	.10	.06	-.01	.42	-	
9. Social T2	.18	.20	.17	.36	.29	.10	.26	.21	-
10. Assertiveness T2	.27	.19	.15	.26	.34	.17	.23	.21	.53

Note. N = 3,876.

Table 4
Latent Intercorrelations of the First-Order Self-Concepts in the TOSCA-10 Study

	1	2	3	4	5	6	7	8	9
1. Math T1	-								
2. Verbal T1	-.31	-							
3. English T1	-.27	.62	-						
4. Economic T1	.14	.19	.07	-					
5. Technical T1	.28	-.25	-.23	.14	-				
6. Math T2	.79	-.26	-.22	.13	.30	-			
7. Verbal T2	-.31	.71	.56	.23	-.21	-.25	-		
8. English T2	-.21	.51	.67	.09	-.15	-.17	.64	-	
9. Economic T2	.16	.13	.12	.47	.00	.31	.29	.19	-
10. Technical T2	.27	-.20	-.19	.08	.78	.33	-.14	-.12	-.01

Note. N = 2,095.

After examining the variances, we were interested in the stabilities of global self-concept modeled by a second-order factor and a composite score (Research Question 3). We examined the temporal stabilities from the three studies across 1 (Study 1), 4 (Study 2), and 7 years (Study 3). Tables 6 and 7 depict the covariances and correlations between the respective time points in the three studies with a stepwise inclusion of self-concepts. Across the three studies and across all sets of self-concepts, the covariances were typically higher for the composite score than for the second-order factor. However, the correlations were almost always higher in the second-order factor, which is most likely a result of the small(er) variances in the second-order factor that boost up the correlations. The differences between the two approaches were small in Study 1 (r_{SOF} = .50 to .70; r_{LCS} = .50 to .59) and more pronounced in Study 2 (r_{SOF} = .50 to .75; r_{LCS} = .39 to .49) and Study 3 (rs_{SOF} = .91; r_{LCS} = .64 to .68). In addition, in Study 3, it was not possible to estimate the correlations of the second-order factor when two or three self-concepts were included due to the negative variances in the second-order factor or nonconvergence (see Table 5). Across Studies 1 and 2, the correlations across the different sets of self-concept were more similar (i.e., they had a smaller range) in the composite score than in the second-order factor.

In sum, we observed mostly low correlations between the lower order self-concepts, which transferred into small variances of the second-order factor (see Table 5). These small variances most likely inflated some of the stabilities in the second-order factor (see Tables 6 and 7). By contrast, the variances of the composite scores were higher and their stabilities were more consistent, given different sets of lower order self-concepts.

External Criteria

As the second major step in our analyses, we were interested in the relations between global self-concept, modeled with the two approaches, with external criteria (Research Question 4). For Studies 1 and 2, we looked at global self-concept's relations to global self-esteem and enjoyment of school. In Study 3, which included only academic self-concepts, we looked at global self-concept's relations to global academic outcomes, namely, an average achievement score, grade point average, and the transition after Grade 10 (academic vs. vocational track).

Table 5
Means and (Explained Proportions of) Variances of the Second-Order Factor and the Composite Score

	Global self-concept: Second-order factor						Global self-concept: Composite score			
	T1			T2			T1		T2	
Set of self-concepts	*M*	$\widehat{\sigma}^2$	$R^{2\,a}$	*M*	$\widehat{\sigma}^2$	$R^{2\,a}$	*M*	$\widehat{\sigma}^2$	*M*	$\widehat{\sigma}^2$
Study 1: BIJU										
M, G	2.09	0.07	.194	2.10	0.11	.224	2.08	0.22	2.11	0.30
M, G, E	2.06	0.07	.187	2.07	0.11	.219	2.07	0.18	2.06	0.23
M, G, E, B	2.00	0.07	.188	2.04	0.11	.230	2.01	0.16	2.02	0.20
M, G, E, B, P	2.02	0.11	.265	2.09	0.13	.273	2.04	0.17	2.07	0.20
S, A	2.28	0.18	.618	2.17	0.23	.765	2.29	0.23	2.16	0.26
S, A, Ap	2.28	0.14	.615	2.20	0.17	.729	2.28	0.17	2.20	0.20
All	2.13	0.08	.302	2.11	0.09	.332	2.13	0.12	2.12	0.14
Study 2: TRAIN										
M, G	2.97	0.07	.261	2.88	0.06	.224	3.00	0.18	2.83	0.20
M, G, E	3.01	0.09	.312	2.87	0.08	.265	3.03	0.16	2.85	0.17
S, A	2.97	0.26	.657	3.04	0.23	.543	3.03	0.33	3.03	0.33
All	3.00	0.13	.379	2.95	0.10	.277	3.03	0.16	2.92	0.15
Study 3: TOSCA-10										
M, V	2.91	-0.12	-	2.87	-0.09	-	2.91	0.15	2.90	0.18
M, V, E			No convergence				2.91	0.15	2.92	0.18
M, V, E, Ec	2.84	0.03	.038[b]	2.82	0.03	.030[b]	2.83	0.12	2.87	0.19
M, V, E, Ec, T	2.80	0.10	.011[b]	2.79	0.10	.011[b]	2.79	0.10	2.83	0.15

Note. $N_{\text{Study 1}}$ = 8,068, $N_{\text{Study 2}}$ = 3,876, $N_{\text{Study 3}}$ = 2,095. M = Math, G = German, E = English, B = Biology, P = Physics, S = Social, A = Assertiveness, Ap = Appearance, V= Verbal, Ec = Economic, T = Technical.
[a] Average R^2across all first-order factors.
[b] For some of the first-order factors, the explained variance could not be computed.

Table 6

Raw Rank-Order Stabilities and Correlations (and Covariances) of the Second-Order Factor and the Composite Score with Self-Esteem and Enjoyment of School in Studies 1 and 2

	Global self-concept: Second-order factor					Global self-concept: Composite score				
	T1		T2		T1 with T2	T1		T2		T1 with T2
Set of self-concepts	*r(cov)* GSC, SE	*r(cov)* GSC, ES	*r(cov)* GSC, SE	*r(cov)* GSC, ES	*r (cov)* GSC_{T1}, GSC_{T2}	*r(cov)* GSC, SE	*r(cov)* GSC, ES	*r(cov)* GSC, SE	*r(cov)* GSC, ES	*r (cov)* GSC_{T1}, GSC_{T2}
Study 1: BIJU										
AC2: M, G	.74 (.12)	.52 (.09)	.78 (.17)	.47 (.08)	.56 (.05)	.42 (.12)	.29 (.09)	.47 (.17)	.29 (.08)	.50 (.13)
AC3: M, G, E	.74 (.12)	.50 (.09)	.76 (.17)	.47 (.08)	.65 (.06)	.49 (.13)	.33 (.09)	.52 (.17)	.33 (.08)	.57 (.12)
AC4: M, G, E, B	.67 (.11)	.44 (.07)	.71 (.16)	.43 (.07)	.70 (.06)	.50 (.12)	.33 (.08)	.55 (.17)	.34 (.08)	.57 (.10)
AC5: M, G, E, B, P	.56 (.12)	.34 (.07)	.63 (.15)	.38 (.07)	.68 (.08)	.49 (.13)	.32 (.08)	.55 (.16)	.34 (.08)	.59 (.11)
NAC2: S, A	.92 (.24)	.34 (.09)	.52 (.14)	.20 (.04)	.54 (.11)	.80 (.24)	.29 (.09)	.48 (.16)	.18 (.05)	.53 (.13)
NAC3: S, A, Ap	1.03(.24)	.31 (.07)	.55 (.16)	.21 (.05)	.50 (.08)	.91 (.23)	.29 (.08)	.51 (.15)	.20 (.04)	.52 (.10)
All	1.01(.18)	.33 (.07)	.66 (.13)	.27 (.05)	.56 (.05)	.77 (.17)	.36 (.08)	.63 (.16)	.34 (.06)	.59 (.08)
Study 2: TRAIN										
AC2: M, G	.61 (.13)	.79 (.13)	.64 (.12)	.77 (.11)	.60 (.04)	.39 (.13)	.46 (.12)	.37 (.13)	.38 (.10)	.45 (.09)
AC3: M, G, E	.50 (.12)	.60 (.11)	.52 (.11)	.57 (.10)	.75 (.06)	.38 (.12)	.44 (.11)	.38 (.12)	.38 (.09)	.49 (.08)
NAC2: S, A	.41 (.17)	.21 (.07)	.50 (.18)	.19 (.05)	.50 (.12)	.35 (.16)	.17 (.06)	.41 (.18)	.15 (.05)	.39 (.13)
All	.45 (.13)	.32 (.07)	.58 (.14)	.34 (.06)	.52 (.06)	.42 (.14)	.35 (.09)	.48 (.14)	.33 (.08)	.47 (.07)

Note. $N_{Study\ 1}$ = 8,068, $N_{Study\ 2}$ = 3,876. AC = Academic self-concept, NAC = Nonacademic self-concept, M = Math, G = German, E = English, B = Biology, P = Physics, S = Social, A = Assertiveness, Ap = Appearance, GSC = Global self-concept, SE = Self-esteem, ES = Enjoyment of school.

Table 7

Raw Rank-Order Stabilities and Correlations (and Covariances) of the Second-Order Factor and the Composite Score with Academic Outcomes in Study 3

	Global self-concept: Second-order factor							Global self-concept: Composite score						
	T1			T2			T1 with T2	T1			T2			T1 with T2
	r(cov) GSC,Ach	*r(cov)* GSC,GPA	*r(cov)* GSC,Tra	*r(cov)* GSC,Ach	*r(cov)* GSC,GPA	*r(cov)* GSC,Tra	*r (cov)* GSC T1, GSC T2	*r(cov)* GSC,Ach	*r(cov)* GSC,GPA	*r(cov)* GSC,Tra	*r(cov)* GSC,Ach	*r(cov)* GSC,GPA	*r(cov)* GSC,Tra	*r (cov)* GSC T1, GSC T2
M, V	- (.06)	- (-.10)	- (.06)	- (.05)	- (-.06)	- (.02)	- (-.11)	.37 (.07)	-.52 (-.13)	.39 (.08)	.32 (.07)	-.31 (-.08)	.12 (.03)	.68 (.11)
M, V, E	No convergence							.32 (.06)	-.42 (-.10)	.36 (.07)	.34 (.07)	-.28 (-.07)	.14 (.03)	.68 (.11)
M, V, E, Ec	.28 (.03)	-.40 (-.05)	.27 (.03)	.25 (.02)	-.32 (-.03)	.15 (.01)	.91[a] (.03)	.33 (.06)	-.40 (-.09)	.34 (.06)	.31 (.07)	-.28 (-.08)	.16 (.03)	.64 (.10)
M, V, E, Ec, T	.16 (.02)	-.25 (-.05)	.19 (.03)	.15 (.02)	-.19 (-.04)	.10 (.02)	.91[a] (.09)	.34 (.05)	-.30 (-.06)	.22 (.04)	.33 (.06)	-.22 (-.05)	.10 (.02)	.66 (.08)

Note. $N_{\text{Study 3}}$ = 2,095. M = Math, V= Verbal, E = English, Ec = Economic, T = Technical, GSC = Global self-concept, Ach = Average achievement score, GPA = grade point average, Tra = Transition after grade 10.

[a] This stability most likely represents math self-concept only because the second-order factor has a large loading on math self-concept and small or negative loadings on the other self-concepts. At the same time, the residual variances and the uniqueness (correlations of the residual variances) of math self-concept are negative.

Table 6 presents the covariances and correlations of global self-concept with global self-esteem and enjoyment of school in Studies 1 and 2. Regarding the relations between global self-concept and self-esteem, across both studies and in all sets of self-concepts, the covariances were very similar in the second-order factor and the composite score. By contrast, the correlations were consistently higher in the second-order factor (Study 1: r_{SOF} = .52 to 1.03; Study 2: r_{SOF} = .41 to .64) than in the composite score (Study 1: r_{LCS} = .42 to .91; Study 2: r_{LCS} = .35 to .48). Across both studies and both time points, reducing the number of self-concepts to model a second-order factor was typically associated with increases in the correlation with self-esteem, whereas this was not the case in the composite score model. For example, the correlations between the second-order factor and global self-esteem were among the highest when only two academic self-concepts were included (Study 1: $r_{SOF} > .74$; Study 2: $r_{SOF} > .61$). In Study 1, some correlations from the second-order factor were implausibly high ($r_{SOF} > 1$).

The covariances and correlations with regard to the second external criterion (i.e., enjoyment of school) in Studies 1 and 2 are also presented in Table 6. Paralleling the findings for self-esteem, the covariances between global self-concept and enjoyment of school were very similar in the second-order factor and the composite score. In line with the findings on self-esteem, most of the correlations were higher in the second-order factor than in the composite score. Across the two studies, the range of the correlations with varying self-concepts was smaller among the correlations from the composite score (Study 1: r_{LCS} = .18 to .36; Study 2: r_{LCS} = .15 to .46) than among the ones from the second-order factor (Study 1: r_{SOF} = .20 to .52; Study 2: r_{SOF} = .19 to .79). In the second-order factor, the correlations were highest when only two or three academic self-concepts were included. Across both studies and both modeling approaches, the correlations between global self-concept and enjoyment of school were higher when global self-concept was based on academic rather than nonacademic self-concepts.

As indicated in Table 7, in Study 3, results on the relations between global self-concept and three academic outcomes (average achievement score, grade point average, decision about the transition) indicated that the covariances and correlations were higher for the composite score (r_{LCS} = |.10| to |.52|) than for the second-order factor (r_{SOF} = |.10| to |.40|). In the second-order factor, it was not possible to estimate the correlations between global self-concept and the academic outcomes when two or three self-concepts were included. Again, this was due to the negative variance of global self-concept in one model and the nonconvergence in the other model (see Table 5).

Overall, the relations to external criteria (i.e., self-esteem, enjoyment of school, academic outcomes) differed between the second-order factor and the composite score. Most importantly, in the second-order factor, some correlations were very high, implausible, or not computable, which was probably the result of the small variances of the second-order factor. By contrast, the results of the composite score were more consistent throughout and largely similar across the different sets of the lower order self-concepts.

Discussion

The present research was aimed at comparing two competing theoretical representations of the Shavelson model (Shavelson et al., 1976): (a) a top-down approach in which global self-concept represents the cause of lower order self-concepts, implemented by a reflective model in terms of a second-order factor model and (b) a bottom-up approach in which lower order self-concepts form a global self-concept, represented by a formative procedure in terms of a model-based latent composite score. Studies prior to this research investigated global self-concept almost exclusively from the perspective of reflective modeling approaches. This development was most likely driven by methodological advances rather than theoretical considerations. Therefore, the present research contrasted the two theoretical ideas and included three independent large-scale studies (Study 1: N = 8,068; Study 2: N = 3,876; Study 3: N = 2,095) in order to empirically evaluate the respective methodological implementations.

Across the three studies (including varying sets of self-concepts and across two time points each), our results provided support for four major findings: First, the correlations between the lower order self-concepts were not consistently high. Particularly in conceptually distinct domains (e.g., math and verbal domains), the correlations were often low or sometimes even negative. Second, global self-concept consistently revealed a higher variance when modeled as a composite score in comparison with a second-order factor. In one case, the second-order factor resulted in a negative variance, and in another case, the second-order factor model did not even converge. Third, the results on the stabilities as well as the results on the relations with external criteria indicated that the composite score typically revealed higher covariances than the second-order factor, whereas for the correlations, it was the opposite, most likely due to the small variances in the second-order factors. Finally, some correlations in the second-order factor were not estimable, implausible, or surprisingly high (correlations of two academic self-concepts and self-esteem) or surprisingly low (e.g., correlations between two academic self-concepts and academic outcomes), whereas this was not the case for the composite score.

The present findings line up with previous studies, which showed that academic lower order self-concepts were barely or even negatively correlated, mirroring the effect described as the internal/external frame of reference effect (e.g., Marsh & Shavelson, 1985; Marsh, 1986b). These low and/or negative correlations contradicted the prerequisites for second-order factor models because second-order factors should be grounded on positive high intercorrelations between indicators (e.g., Bollen & Lennox, 1991). The small variances in the second-order factors, which we found across all three studies, are likely the result of these low intercorrelations. Squeezing small correlations across lower order self-concepts into a common factor resulted in a variance-restricted second-order factor, which reproduced only small overall differences between individuals. By contrast, the composite score did not depend on substantial intercorrelations between lower order self-concepts. The low variances in the second-order factors fired back when we used the respective global self-concept for further analysis: The small variances dramatically influenced all standardized coefficients because of division by the variance, such as when looking at stabilities or correlations with external constructs. This could be problematic as it leads researchers to believe that there are high correlations when they are just artificially inflated.

When reviewing the strengths of the correlations with external criteria, we observed that the coefficients showed various incompatibilities with theoretical considerations. Theoretically, we would expect that global self-concept and global self-esteem would be more strongly correlated when considering a broader compared with a smaller spectrum of academic self-concepts. However, given the present findings on the second-order factor, correlations with self-esteem increase when less academic self-concepts are included. More specifically, in Studies 1 and 2, when only math and German self-concept were included, correlations between the second-order factor and self-esteem were surprisingly high (r = .61 to .78). This could indicate that students' general feelings of worth are to a large extent determined by how they perceive their abilities in math and German, which is questionable given previous findings on the relation between academic self-concept and self-esteem over time (e.g., Marsh & O'Mara, 2008; Trautwein et al., 2006). Over and above the correlations that contradicted previous theoretical and empirical findings, some of the correlations in the second-order factor were factually not computable or implausible. More specifically, in Study 3, the negative variance of the second-order factor, which was an artificial result of the negative intercorrelations between lower order self-concepts, made further analysis impossible. In Study 1, two correlations exceeded a value

of 1, which was another finding that fell outside the plausible range. Interestingly, these correlations appeared not only when the set of self-concepts were barely correlated but also when there were substantial correlations between the included self-concepts.

Overall, when modeling global self-concept by a composite score, we revealed a more consistent and theoretically plausible pattern of results. More specifically, the relations with external criteria remained largely similar when additional self-concepts were included, there were no implausible correlations, and the sizes and patterns of the correlations were more in line with theoretical considerations and previous empirical findings on the self-concepts' relations with self-esteem, enjoyment of school, and academic outcomes (Marsh & O'Neill, 1984; Marsh & Yeung, 1998). For example, in Studies 1 and 2, when only math and German self-concepts were included, the correlations with self-esteem were medium in size (r = .37 to .47).

A Latent Composite Score Approach: A Fresh Start for the Shavelson Model?

Research on the structure of self-concept has a vivid history, beginning in a dustbowl and subsequently sparked by Shavelson's, Marsh's, and Brunner's theoretical and empirical specifications (e.g., Brunner et al., 2010; Marsh & Shavelson, 1985; Marsh, 1990; Shavelson et al., 1976). These developments have answered multiple questions and have led to important gains in the understanding of self-concept. Transferring concepts applied to the hierarchical structure of cognitive abilities, previous research on the hierarchy of self-concept has embraced the framework of confirmatory factor analysis. This analytical approach follows a top-down logic by which higher order construct cause lower order constructs. At the same time, theoretical ideas instead suggest a bottom-up process by which lower order self-concepts form higher order self-concepts. One reason why this theoretical idea has not garnered much empirical interest is that statistical procedures for modeling bottom-up ideas (i.e., formative approaches) have not been very sophisticated. A method that was recently developed—the model-based latent composite score (Rose et al., 2019)—was responsible for the present turnaround.

Overall, three insights from our analyses have convinced us to favor the composite score model over the second-order factor model: (a) The composite score consistently revealed higher variances of global self-concept, (b) the composite score yielded more consistent correlations, given varying sets of self-concepts, and (c) the composite score resulted in a more plausible pattern of correlations with respect to their size and theoretical reasonableness. By contrast, second-order factor models resulted in lower variances of global self-concept that

boosted correlations to surprisingly high and sometimes clearly implausible and impermissibly high values.

On the basis of the empirical advantages offered by the latent composite score model, we conclude that a bottom-up model represents global self-concept better than a top-down model does. Hence, according to the present research, a latent composite score modeling approach could provide a fresh start in research on the structure of self-concept. A latent composite score approach suggests that global self-concept is formed by a lower order self-concept. This approach incorporates the assumptions of the multidimensional, hierarchical Shavelson model with global self-concept at the apex of the hierarchy (Shavelson et al., 1976). At the same time, it is in line with findings from the Marsh/Shavelson model and the internal/external frame of reference model (Marsh & Shavelson, 1985; Marsh, 1986b, 1987, 1990; Shavelson & Marsh, 1986) because the composite score approach does not call for substantial correlations between lower order self-concepts, such as math and verbal self-concept. Furthermore, the latent composite score approach is in line with a central characteristic of the Nested Marsh/Shavelson model (Brunner et al., 2009; Brunner et al., 2010) in that it incorporates global (academic) self-concept. Hence, a composite score approach unifies previous assumptions on the structure of self-concept by reframing the Shavelson model as a formation process. The present research was aimed at providing a starting point, which needs to be followed by a more granulated understanding of this formation process.

Beyond Self-Concept: Theoretical Assumptions and Statistical Models

Although our study focused on self-concept, we believe that our approach might be of more general interest in educational psychology. Generally speaking, theoretical considerations about the underlying processes of constructs and research questions need to be more strongly connected to the application of statistical models (Borsboom et al., 2003; Rhemtulla et al., 2015, 2019). Whenever one expects a formation process, the latent variable in reflective models may not be appropriate to represent the construct of interest. In addition, even if researchers theoretically assume reflective processes, empirical results (e.g., low intercorrelations among lower order constructs) can challenge these reflective ideas. A conservative consequence could be that researchers end up refraining from modeling a higher order factor, such as in previous research on self-concept (e.g., Marsh, 1990; Rentzsch, Wenzler, & Schütz, 2016). However, by contrast, the higher order factor might also be formed by the lower order

factors instead of being the cause of the lower order factors. Thus, a different modeling approach would be needed.

In educational psychology, reflective modeling approaches are the default procedure for modeling relations between higher and lower order constructs. In other fields of research, however, formative approaches for multidimensional hierarchical constructs are well established, for example, in management constructs such as strategy or job satisfaction (Law, Wong, & Mobley, 1998; Podsakoff, Shen, & Podsakoff, 2004). Formative approaches for multidimensional constructs might also be embraced more instensively by future psychological research. This needs to start by giving deeper consideration to the ontology of multidimensional constructs. In addition, when there are low correlations between lower order dimensions, this strongly suggests formative considerations. In this regard, even the facade of the psychological pioneer research field in confirmatory factor analysis, the field of intellectual abilities, recently began to crumble with regard to their long-standing tradition of using reflective higher order models (Conway & Kovacs, 2015; Eid, Geiser, Koch, & Heene, 2017). In educational psychology, the relevance of formative approaches might be particularly important, given the wide range of multidimensional, hierarchical constructs in this field. On the one hand, this could apply to global constructs that arise from perceptions across different school subjects, which also face dimensional comparison processes, such as interest, effort, or students' perceptions of teaching quality (e.g., Jaekel, Göllner, & Trautwein, in press). On the other hand, more globally speaking, this could apply to global constructs that are not unidimensional due to the specificity of their components such as task values (Gaspard et al., 2015; Gaspard, Häfner, Parrisius, Trautwein, & Nagengast, 2017), or, in line with research on job satisfaction, this could apply to students' school satisfaction.

Limitations and Future Directions

We analyzed the consequences of top-down and bottom-up representations of the Shavelson model in a comprehensive research framework using multiple studies, multiple sets of self-concepts, and two time points each. This research was aimed at providing a starting point for rethinking the global self-concept. However, there are some limitations that need to be taken into account.

First, it is important to consider the selection of self-concepts. In line with the Shavelson model, we have covered both academic and nonacademic self-concepts. However, it is important to note that the self-concept domains used in this research are by no means exhaustive.

Shavelson and colleagues already emphasized that their illustration of the structure of self-concept is just one possible representation of the universe of self-concepts (Shavelson et al., 1976). In contrast to reflective models, which would assume that indicators are exchangeable because they are caused by one common factor, the selection of indicators is decisive in a formative modeling procedure (Bollen & Lennox, 1991). Here, the selection of self-concepts constitutes global self-concept, and therefore, the composite score needs to be interpreted in the wake of its constitutional elements. For example, when only academic self-concepts are included, the composite score reflects global academic self-concept, whereas when only non-academic self-concepts are included, the composite score represents global nonacademic self-concept. In the present study, the results of the relations between the composite score and external criteria differed in particular with regard to the inclusion of nonacademic self-concepts. For example, global self-esteem seemed to be most strongly related to a global self-concept whenever we included nonacademic self-concepts, which is in line with previous findings (e.g., Donnellan, Trzesniewski, Conger, & Conger, 2007; Harter, 2003). At the same time, correlations were largely similar across different sets of academic self-concepts. However, future research using different sets of self-concepts needs to test the robustness of this pattern of results.

Second, in our application of the model-based latent composite score, we gave all self-concepts equal weights, which is a widely used and justifiable procedure (Wainer, 1976). Yet, by contrast, in the second-order factor model, weights (i.e., the factor loadings of the second-order factor) were estimated, and hence, they differed across the lower order self-concepts. In the composite score model, these weights needed to be assigned in advance, and therefore, they needed to be a result of theoretical considerations. For example, if we want to approximate the relation between global self-concept and global self-esteem, nonacademic self-concepts need to be given more weight than academic self-concepts. Another stream of research proposed to extend the Shavelson model by considering the individual importance of the lower order self-concepts (James, 1890/1963). Despite the fact that empirical studies have failed to find support for this model using second-order factor models (Marsh, 1986a; Marsh & Scalas, 2018; Scalas, Marsh, Nagengast, & Morin, 2013), this could be a direction for future research using a latent composite score approach.

Finally, previous research has argued that top-down models should be tested against bottom-up models by using multiwave-multivariable structural equation modeling, such as cross-lagged panel models (Harris et al., 2018; Marsh & O'Mara, 2008; Rentzsch & Schröder-Abé, 2018; Trautwein et al., 2006). This field of research has tested ideas that are conceptually

similar to those from the present study with temporal predictions across yearly waves of assessment. In contrast to the present study, these studies (a) focused on processes over time rather than the directional link *within* the hierarchy of self-concept and (b) used unidimensional global self-esteem as an indicator of global self-concept. However, it is unclear whether this measure provides the optimal operationalization of the apex of the Shavelson hierarchy. Future research could merge our approach with the longitudinal predictions in order to understand how lower order self-concepts and global self-concept, modeled as a composite score, as well as unidimensional global self-esteem are related over time. This could enrich our understanding about not only the prospective relations between lower and higher order self-concepts but also the similarities and differences between global self-concept and global self-esteem.

Conclusion

The present research was aimed at integrating theoretical considerations and statistical modeling approaches in the implementation of global self-concept. We proposed that the previous focus on conventional second-order factor models be changed to a focus on model-based latent composite scores. Over and above research on self-concept, the present findings should pave the way for turnarounds in the analyses of multidimensional, hierarchical constructs: Theoretical considerations of bottom-up processes combined with their empirical foundations (e.g., low correlations between the lower order dimensions), should lead researchers to implementations that use formative modeling approaches (e.g., the model-based latent composite score). According to our research, for students such as Jasmin and Rafael, instead of their domain-specific self-concepts being determined by their global self-concept, these students rather aggregate their self-concepts in different areas of their lives in order to generate a global self-concept. A more granulated understanding of Jasmin's and Rafael's self-concept formation processes needs to build the foundation of future self-concept research.

References

Baumert, J., Roeder, P. M., Gruehn S., Heyn S., Köller, O., Rimmele, R.,. . . Seipp, B. (1996). Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU) [Learning processes, educational careers, and psychosocial development in adolescence]. In K. P. Treumann, G. Neubauer, R. Möller, & J. Abel (Eds.), *Methoden und Anwendungen empirischer pädagogischer Forschung* (pp. 170–180). Münster: Waxmann.

Blalock, H. M. (1964). *Causal inferences in nonexperimental research*. Chapel Hill: University of North Carolina Press.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*(2), 305–314. https://doi.org/10.1037/0033-2909.110.2.305

Bollen, K. A., & Bauldry, S. (2011). Three Cs in measurement models: causal indicators, composite indicators, and covariates. *Psychological Methods*, *16*(3), 265–284. https://doi.org/10.1037/a0024448

Bollen, K. A., & Diamantopoulos, A. (2017). In Defense of Causal-Formative Indicators: A Minority Report. *Psychological Methods*, *22*(3), 581–596. https://doi.org/10.1037/met0000056

Borsboom, D. (2008). Latent Variable Theory. *Measurement: Interdisciplinary Research and Perspectives*, *6*(1-2), 25–53. https://doi.org/10.1080/15366360802035497

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*(2), 203–219. https://doi.org/10.1037/0033-295X.110.2.203

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Brunner, M., Keller, U., Dierendonck, C., Reichert, M., Ugen, S., Fischbach, A., & Martin, R. (2010). The structure of academic self-concepts revisited: The nested Marsh/Shavelson model. *Journal of Educational Psychology*, *102*(4), 964–981. https://doi.org/10.1037/a0019644

Brunner, M., Keller, U., Hornung, C., Reichert, M., & Martin, R. (2009). The cross-cultural generalizability of a new structural model of academic self-concepts. *Learning and Individual Differences*, *19*(4), 387–403. https://doi.org/10.1016/j.lindif.2008.11.008

Brunner, M., Lüdtke, O., & Trautwein, U. (2008). The internal/external frame of reference model revisited: Incorporating general cognitive ability and general academic Self-Concept. *Multivariate Behavioral Research*, *43*(1), 137–172. https://doi.org/10.1080/00273170701836737

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multi-trait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. https://doi.org/10.1037/h0046016

Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, *12*(4), 381–398. https://doi.org/10.1037/1082-989X.12.4.381

Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods*, *19*(2), 300–315. https://doi.org/10.1037/a0033805

Conway, A. R. A., & Kovacs, K. (2015). New and emerging models of human intelligence. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(5), 419–426. https://doi.org/10.1002/wcs.1356

Coopersmith, S. (1967). *The antecedents of self-esteem*. San Francisco: W. H. Freeman.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, *61*(12), 1203–1218. https://doi.org/10.1016/j.jbusres.2008.01.009

Donnellan, M. B., Trzesniewski, K. H., Conger, K. J., & Conger, R. D. (2007). A three-wave longitudinal study of self-evaluations during young adulthood. *Journal of Research in Personality*, *41*(2), 453–472. https://doi.org/10.1016/j.jrp.2006.06.004

Donnellan, M. B., Trzesniewski, K. H., & Robins, R. W. (2011). Self-esteem: Enduring issues and controversies. In T. Chamorro-Premuzic (Ed.), *The Wiley-Blackwell handbook of individual differences* (pp. 718–746). Chichester: Wiley-Blackwell.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155–174. https://doi.org/10.1037/1082-989X.5.2.155

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods*, *22*(3), 541–562. https://doi.org/10.1037/met0000083

Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C(M-1) model. *Psychological Methods*, *8*(1), 38–60. https://doi.org/10.1037/1082-989X.8.1.38

Fend, H., & Prieser, H.-G. (Eds.). (1986). *Dokumentation der Skalen des Projekts "Entwicklung im Jugendalter" [Documentation of the instruments of the project "Development in adolescence"]*. Konstanz: Sozialwissenschaftliche Fakultät der Universität Konstanz.

Fitts, W. H. (1965). *A manual for the Tennessee Self-Concept Scale*. Nashville, TN: Counselor Recordings and Tests.

Gaspard, H., Dicke, A.-L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., & Nagengast, B. (2015). More value through greater differentiation: Gender differences in value beliefs about math. *Journal of Educational Psychology*, *107*(3), 663–677. https://doi.org/10.1037/edu0000003

Gaspard, H., Häfner, I., Parrisius, C., Trautwein, U., & Nagengast, B. (2017). Assessing task values in five subjects during secondary school: Measurement structure and mean level differences across grade level, gender, and academic subject. *Contemporary Educational Psychology*, *48*, 67–84. https://doi.org/10.1016/j.cedpsych.2016.09.003

Greenwald, A. G. (2012). There Is Nothing So Theoretical as a Good Method. *Perspectives on Psychological Science*, *7*(2), 99–108. https://doi.org/10.1177/1745691611434210

Harris, M. A., Wetzel, E., Robins, R. W., Donnellan, M. B., & Trzesniewski, K. H. (2018). The development of global and domain self-esteem from ages 10 to 16 for Mexican-origin youth. *International Journal of Behavioral Development*, *42*(1), 4–16. https://doi.org/10.1177/0165025416679744

Harter, S. (2003). The development of self-representations during childhood and adolescence. In In M. R. Leary & J. P. Tangney (Eds.), *Handbook of Self and Identity* (pp. 610–642). New York: Wiley.

Hishiki, P. C. (1969). Self-concepts of 6th-grade girls of Mexican-American descent. *California Journal of Educational Research*. (20), 56–62.

Jaekel, A., Göllner, R., & Trautwein, U. (in press). How Students' Perceptions of Teaching Quality in one Subject are Impacted by the Grades They Receive in Another Subject — Dimensional Comparisons in Student Evaluations of Teaching Quality. *Journal of Educational Psychology*.

James, W. (1890/1963). *The principles of psychology*. New York: Holt, Rinehart & Winston.

Jerusalem, M. (1984). *Selbstbezogene Kognitionen in schulischen Bezugsgruppen: Eine Längsschnittstudie [Self-related cognitions in school contexts]*. Berlin: Freie Universität Berlin, Institut für Psychologie.

Jonkmann, K., Rose, N., & Trautwein, U. (Eds.). (2013). *Tradtition und Innovation: Entwicklungsverläufe an Haupt- und Realschulen in Baden-Württemberg und Mittelschulen in Sachsen - Abschlussbericht für die Länder Baden-Württemberg und Sachsen [Tradition and innovation: Academic and psychosocial development in vocational track schools in the states of Baden-Württemberg and Sachsen]*. Tübingen, Germany: Unpublished project report, University of Tübingen.

Jopt, U.-J. (1978). *Selbstkonzept und Ursachenerklärung in der Schule: Zur Attribuierung von Schulleistungen [Self-concept and explanation of causes in school: On the attribution of school performance]*. Bochum: Kamp.

Law, K. S., Wong, C.-S., & Mobley, W. M. (1998). Toward A Taxonomy of Multidimensional Constructs. *Academy of Management Review*, *23*(4), 741–755. https://doi.org/10.5465/amr.1998.1255636

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(1), 59–72. https://doi.org/10.1207/s15328007sem1301_3

Ludwig, D. J., & Maehr, M. L. (1967). Changes in self concept and stated behavioral preferences. *Child Development*, *38*(2), 453. https://doi.org/10.2307/1127302

MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. (2005). The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology, 90*(4), 710–730. https://doi.org/10.1037/0021-9010.90.4.710

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

Marsh, H. W. (1986b). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, *23*(1), 129–149. https://doi.org/10.3102/00028312023001129

Marsh, H. W. (1987). The hierarchical structure of self-concept and the application of hierarchical confirmatory factor analysis. *Journal of Educational Measurement*, *24*(1), 17–39. https://doi.org/10.1111/j.1745-3984.1987.tb00259.x

Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early adulthood. *Journal of Educational Psychology*, *81*(3), 417–430. https://doi.org/10.1037/0022-0663.81.3.417

Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology*, *82*(4), 623–636. https://doi.org/10.1037/0022-0663.82.4.623

Marsh, H. W., & Ayotte, V. (2003). Do Multiple Dimensions of Self-Concept Become More Differentiated With Age? The Differential Distinctiveness Hypothesis. *Journal of Educational Psychology*, *95*(4), 687–706. https://doi.org/10.1037/0022-0663.95.4.687

Marsh, H. W., & Craven, R. (1997). Academic self-concept. In G. D. Phye (Ed.), *Handbook of classroom assessment. Learning, achievement, and adjustment* (pp. 131–198). Orlando, FL: Academic Press. https://doi.org/10.1016/B978-012554155-8/50008-9

Marsh, H. W., Craven, R. G., & Debus, R. (1991). Self-concepts of young children 5 to 8 years of age: Measurement and multidimensional structure. *Journal of Educational Psychology*, *83*(3), 377–392. https://doi.org/10.1037/0022-0663.83.3.377

Marsh, H. W., & Hattie, J. A. (1996). Theoretical models in self-concept. In B. Bracken (Ed.), *Handbook of self-concept* (pp. 38–90). Hillsdale, NJ: Erlbaum.

Marsh, H. W., & Hau, K.-T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology*, *96*(1), 56–67. https://doi.org/10.1037/0022-0663.96.1.56

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*(1), 151–170. https://doi.org/10.1016/j.cedpsych.2006.10.008

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. *Psychological Bulletin*, *97*(3), 562–582. https://doi.org/10.1037/0033-2909.97.3.562

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

Marsh, H. W., & O'Neill, R. (1984). Self Description Questionnaire III: The Construct Validity Of Multidimensional Self-Concept Ratings By Late Adolescents. *Journal of Educational Measurement*, *21*(2), 153–174. https://doi.org/10.1111/j.1745-3984.1984.tb00227.x

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: A Multidisciplinary Journal*, *25*(1), 137–159. https://doi.org/10.1080/10705511.2017.1370377

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, *20*(3), 107–123. https://doi.org/10.1207/s15326985ep2003_1

Marsh, H. W., & Yeung, A. S. (1998). Longitudinal structural equation models of academic self-concept and achievement: Gender differences in the development of math and English constructs. *American Educational Research Journal*, *35*(4), 705–738. https://doi.org/10.3102/00028312035004705

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, *22*(1), 114–140. https://doi.org/10.1037/met0000078

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. https://doi.org/10.1007/BF02294825

Muthen, B. O., & Satorra, A. (1995). Complex Sample Data in Structural Equation Modeling. *Sociological Methodology*, *25*, 267. https://doi.org/10.2307/271070

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide (8th ed.)*. Los Angeles, CA: Muthén & Muthén.

Podsakoff, N. P., Shen, W., & Podsakoff, P. M. (2004). The role of formative measurement models in strategic management research: Review, critique, and implications for future research. In D. J. Ketchen & D. D. Bergh (Eds.), *Research methodology in strategy and management* (pp. 197–252). Amsterdam, Boston: Elsevier. https://doi.org/10.1016/S1479-8387(06)03008-6

Ravens-Sieberer, U., Gosch, A., Abel, T., Auquier, P., Bellach, B.-M., Bruil, J.,. . . Rajmil, L. (2001). Quality of life in children and adolescents: A European public health perspective. *Sozial- und Präventivmedizin SPM*, *46*(5), 294–302. https://doi.org/10.1007/BF01321080

Rentzsch, K., & Schröder-Abé, M. (2018). Stability and change in domain-specific self-esteem and global self-esteem. *European Journal of Personality*, *32*(4), 353–370. https://doi.org/10.1002/per.2167

Rentzsch, K., Wenzler, M. P., & Schütz, A. (2016). The structure of multidimensional self-esteem across age and gender. *Personality and Individual Differences*, *88*, 139–147. https://doi.org/10.1016/j.paid.2015.09.012

Rhemtulla, M., van Bork, R., & Borsboom, D. (2015). Calling models with causal indicators "measurement models" implies more than they can deliver. *Measurement: Interdisciplinary Research and Perspectives*, *13*(1), 59–62. https://doi.org/10.1080/15366367.2015.1016343

Rhemtulla, M., van Bork, R., & Borsboom, D. (2019). Worse than measurement error: Consequences of inappropriate latent variable measurement models. *Psychological Methods.* Advance online publication. https://doi.org/10.1037/met0000220

Rose, N., Wagner, W., Mayer, A., & Nagengast, B. (2019). Model-based manifest and latent composite scores in structural equation models. *Collabra: Psychology*, *5*(1), 9. https://doi.org/10.1525/collabra.143

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Scalas, L. F., Marsh, H. W., Nagengast, B., & Morin, A. J. S. (2013). Latent-variable approaches to the Jamesian model of importance-weighted averages. *Personality and Social Psychology Bulletin*, *39*(1), 100–114. https://doi.org/10.1177/0146167212465321

Sears, P. S., Adenubi, M., Block, M., Crist, J., Gamble, J., & Hubner, J. J. (1972). *Effective reinforcement for achievement behaviors in disadvantaged children: The first year. Technical Report No. 30*. Stanford: Stanford Center for Research and Development in Teaching, 1972. (ERIC Document Reproduction Service No. ED 067 442).

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, *46*(3), 407–441. https://doi.org/10.3102/00346543046003407

Shavelson, R. J., & Marsh, H. W. (1986). On the structure of self-concept. In R. Schwarzer (Ed.), *Anxiety and cognition* (pp. 305–330). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Soares, A. T., & Soares, L. M. (1969). Self-perceptions of culturally disadvantaged children. *American Educational Research Journal*, *6*(1), 31–45. https://doi.org/10.3102/00028312006001031

Soares, L. M., & Soares, A. T. (1977). The self-concept: Mini, maxi, multi ? *Paper presented at the Annual Meeting of the American Educational Research Association, New York.*

Swann, W. B., Chang-Schneider, C., & Larsen McClarty, K. (2007). Do people's self-views matter? Self-concept and self-esteem in everyday life. *American Psychologist*, *62*(2), 84–94. https://doi.org/10.1037/0003-066X.62.2.84

Trautwein, U. (2003). *Schule und Selbstwert [Schools and self-esteem]*. Münster: Waxmann.

Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Self-esteem, academic self-concept, and achievement: how the learning environment moderates the dynamics of self-concept. *Journal of Personality and Social Psychology*, *90*(2), 334–349. https://doi.org/10.1037/0022-3514.90.2.334

Trautwein, U., Nagy, G., & Maaz, K. (2011). Soziale Disparitäten und die Öffnung des Sekundarschulsystems [Social disparities and the opening of the secondary school system in Germany - A study of the transition from lower to upper secondary education]. *Zeitschrift für Erziehungswissenschaft*, *14*(3), 445–463. https://doi.org/10.1007/s11618-011-0220-5

Vernon, P. E. (1950). *The structure of human abilities*. New York: Wiley.

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin*, *83*(2), 213–217. https://doi.org/10.1037/0033-2909.83.2.213

Zirkel, P. A. (1971). Self-concept and the “disadvantage” of ethnic group membership and mixture. *Review of Educational Research*, *41*(3), 211–225. https://doi.org/10.3102/00346543041003211

Appendix

Table A1
Fit Indices for the Longitudinal Models Tested in the BIJU Study

	χ^2	*df*	SF	CFI	TLI	RMSEA	SRMR
Correlated first-order factors	9581.04	2328	1.205	0.943	0.938	0.02	0.033
Latent composite score							
LCS All	9581.03	2328	1.205	0.943	0.938	0.020	0.033
LCS AC5	9581.03	2328	1.205	0.943	0.938	0.020	0.033
LCS AC4	9581.03	2328	1.205	0.943	0.938	0.020	0.033
LCS AC3	9581.03	2328	1.205	0.943	0.938	0.020	0.033
LCS AC2	9581.03	2328	1.205	0.943	0.938	0.020	0.033
LCS NAC3	9581.03	2328	1.205	0.943	0.938	0.020	0.033
LCS NAC2	9581.03	2328	1.205	0.943	0.938	0.020	0.033
Latent composite score + external criteria							
LCS All + Ex	12838.99	3344	1.195	0.935	0.930	0.019	0.034
LCS AC5 + Ex	12838.99	3344	1.195	0.935	0.930	0.019	0.034
LCS AC4 + Ex	12838.99	3344	1.195	0.935	0.930	0.019	0.034
LCS AC3 + Ex	12838.99	3344	1.195	0.935	0.930	0.019	0.034
LCS AC2 + Ex	12838.99	3344	1.195	0.935	0.930	0.019	0.034
LCS NAC3 + Ex	12838.99	3344	1.195	0.935	0.930	0.019	0.034
LCS NAC2 + Ex	12838.99	3344	1.195	0.935	0.930	0.019	0.034
Second-order factor							
SOF All	12775.12	2437	1.208	0.919	0.915	0.023	0.066
SOF AC5	10866.98	2416	1.207	0.934	0.903	0.021	0.046
SOF AC4	10733.98	2401	1.207	0.935	0.931	0.021	0.044
SOF AC3	10630.38	2382	1.206	0.935	0.931	0.021	0.041
SOF AC2	10591.85	2360	1.207	0.935	0.930	0.021	0.040
SOF NAC3	10372.55	2382	1.205	0.937	0.933	0.020	0.037
SOF NAC2	10133.87	2360	1.205	0.939	0.934	0.020	0.035

Second-order factor + external criteria							
SOF All + Ex	16807.30	3509	1.197	0.910	0.906	0.022	0.062
SOF AC5 + Ex	14394.25	3464	1.197	0.926	0.922	0.020	0.044
SOF AC4 + Ex	14138.99	3441	1.197	0.927	0.923	0.020	0.042
SOF AC3 + Ex	13960.16	3414	1.196	0.928	0.923	0.020	0.040
SOF AC2 + Ex	13876.83	3384	1.196	0.929	0.923	0.020	0.038
SOF NAC3 + Ex	15041.34	3420	1.195	0.921	0.916	0.021	0.038
SOF NAC2 + Ex	13420.54	3384	1.196	0.932	0.926	0.019	0.035

Note. N = 8,068. Results based on the MLR estimator (Muthén, & Muthén, 1998-2017); SF = Scaling Factor; CFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.

Table A2
Fit Indices for the Longitudinal Models Tested in the TRAIN Study

	χ^2	*df*	SF	CFI	TLI	RMSEA	SRMR
Correlated first-order factors	1702.73	540	1.132	0.950	0.935	0.027	0.046
Latent composite score							
LCS All	1702.73	540	1.132	0.950	0.935	0.027	0.046
LCS AC3	1702.73	540	1.132	0.950	0.935	0.027	0.046
LCS AC2	1702.73	540	1.132	0.950	0.935	0.027	0.046
LCS NAC2	1702.73	540	1.132	0.950	0.935	0.027	0.046
Composite score + external criteria							
LCS All + Ex	4458.74	1216	1.125	0.903	0.886	0.029	0.048
LCS AC3 + Ex	4458.74	1216	1.125	0.903	0.886	0.029	0.048
LCS AC2 + Ex	4458.74	1216	1.125	0.903	0.886	0.029	0.048
LCS NAC2 + Ex	4458.74	1216	1.125	0.903	0.886	0.029	0.048
Second-order factor							
SO All	2286.69	577	1.138	0.926	0.910	0.031	0.058
SO AC3	1993.78	567	1.135	0.939	0.924	0.029	0.056
SO AC2	2069.43	557	1.133	0.935	0.918	0.030	0.057
SO NAC2	2001.99	557	1.134	0.938	0.921	0.029	0.054
Second-order factor + external criteria							
SOF All + Ex	5341.36	1285	1.1278	0.879	0.865	0.032	0.062
SOF AC3 + Ex	4966.64	1260	1.1249	0.889	0.874	0.031	0.057
SOF AC2 + Ex	4863.95	1241	1.1254	0.892	0.875	0.031	0.055
SOF NAC2 + Ex	4787.45	1242	1.1272	0.894	0.878	0.030	0.052

Note. N = 3,876. Results based on the MLR estimator (Muthén, & Muthén, 1998-2017); SF = Scaling Factor; CFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.

Table A3
Fit Indices for the Longitudinal Models Tested in the TOSCA-10 Study

	χ^2	*df*	SF	CFI	TLI	RMSEA	SRMR
Correlated first-order factors	3439.16	941	1.089	0.935	0.922	0.036	0.048
Latent composite score							
LCS All	3439.16	941	1.089	0.935	0.922	0.036	0.048
LCS AC4	3439.16	941	1.089	0.935	0.922	0.036	0.048
LCS AC3	3439.16	941	1.089	0.935	0.922	0.036	0.048
LCS AC2	3439.16	941	1.089	0.935	0.922	0.036	0.048
Latent composite score + external criteria							
LCS All + Ex	3798.11	1093	1.077	0.939	0.926	0.034	0.046
LCS AC4 + Ex	3798.11	1093	1.077	0.939	0.926	0.034	0.046
LCS AC3 + Ex	3798.11	1093	1.077	0.939	0.926	0.034	0.046
LCS AC2 + Ex	3798.11	1093	1.077	0.939	0.926	0.034	0.046
Second-order factor							
SOF All	3805.77	978	1.091	0.926	0.915	0.037	0.086
SOF AC4	3783.21	975	1.091	0.927	0.915	0.037	0.085
SOF AC3	3720.98	968	1.091	0.928	0.916	0.037	0.084
SOF AC2	4120.40	958	1.091	0.917	0.903	0.040	0.122
Second-order factor + external criteria							
SOF All + Ex	5123.64	1162	1.077	0.911	0.898	0.040	0.110
SOF AC4 + Ex	5001.96	1151	1.076	0.913	0.900	0.040	0.102
SOF AC3 + Ex			No convergence				
SOF AC2 + Ex	4875.48	1118	1.079	0.915	0.899	0.040	0.115

Note. N = 2,095. Results based on the MLR estimator (Muthén, & Muthén, 1998-2017); SF = Scaling Factor; CFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.

4 STUDY 2: HOW STATE AND TRAIT VERSIONS OF SELF-ESTEEM AND DEPRESSIVE SYMPTOMS AFFECT THEIR INTERPLAY: A LONGITUDINAL EXPERIMENTAL EXAMINATION

Braun, L., Göllner, R., Rieger, S., Trautwein, U., & Spengler, M. (2020). How state and trait versions of self-esteem and depressive symptoms affect their interplay: A longitudinal experimental investigation. *Journal of Personality and Social Psychology*. Advance online publication. https://doi.org/10.1037/pspp0000295

Abstract

States and traits are important concepts in psychological research. They can be operationalized (a) by using measures that employ different time frames and (b) by applying statistical models that decompose the variance. However, the effects of using variations in states and traits by applying measurement and modeling approaches have yet to be merged and studied systematically. The present study addressed this topic by revisiting an intensively studied research question, namely: What is the longitudinal relation between self-esteem and depressive symptoms? To do so, we created state and trait versions of questionnaires by systematically changing the time frame ("during the last 2 weeks" vs. "in general") that was used to measure self-esteem and depressive symptoms and in addition, by using state-trait statistical models. We conducted an exploratory study (N = 683) and a confirmatory replication study (N = 1,087) with samples of university students, designed as a 2 × 2 longitudinal experiment with four time points spanning 1 semester. Our results indicated that first, consistently across the two studies, trait time frames revealed higher proportions of trait variance than state time frames. Second, across the two studies, the well-researched vulnerability effect, which postulates that low self-esteem predicts depressive symptoms, only held when trait time frames for self-esteem were applied and traditional cross-lagged models were used. Third, when controlling for stable trait differences, cross-lagged results were least consistent when trait time frames were used, which highlighted the interdependency involved in measuring and modeling states and traits.

Keywords: states and traits, self-esteem, depressive symptoms, cross-lagged panel model (CLPM), latent state-trait cross-lagged panel model (LST-CLPM)

How State and Trait Versions of Self-Esteem and Depressive Symptoms Affect Their Interplay: A Longitudinal Experimental Investigation

When studying psychological constructs and their patterns of relations, researchers face important decisions about how to measure and model their constructs of interest. In this context, an important differentiation concerning measurement and modeling involves the conceptualization of states and traits. States and traits can be operationalized (a) by using measurements that employ the different time frames in questionnaires (e.g., "In general" vs. "During the last 2 weeks") and (b) by applying certain statistical models that decompose the variance (e.g., Kenny & Zautra, 1995, 2001; Steyer, Mayer, Geiser, & Cole, 2015; Steyer, Schmitt, & Eid, 1999). So far, even for research questions that have piqued the interest of large numbers of researchers, the interdependencies involved in the measuring and modeling of states and traits have yet to be analyzed systematically. With this research, we aimed to address this research gap and investigate the combined effect of measuring and modeling states and traits in the context of an intensively studied research question in the field of personality psychology: What is the longitudinal interplay between self-esteem and depressive symptoms? In fact, this research question offers a particularly interesting context for our research because previous studies have relied on measures of self-esteem as a trait (i.e., "In general …") and depressive symptoms as a state (i.e., "During the last 2 weeks…") and used statistical models that did not disentangle states and traits (for an exception, see Masselink et al., 2018). In the present study, we used an experimental, longitudinal design that included two different approaches for measuring self-esteem and depressive symptoms (State: "During the last 2 weeks"; Trait: "In general…"). We combined this approach with the use of statistical models that are able to disentangle states and traits. To increase the robustness of our results, we conducted two independent studies: an exploratory study and a preregistered confirmatory replication study.

States and Traits

The differentiation between states and traits and their relevance for individual behavior has been debated for decades (e.g., Davitz, 1969; Eysenck, 1983; Fleeson & Jayawickreme, 2015; McCrae & Costa, 2008; Nezlek, 2007; Roberts, 2018; Roberts & Jackson, 2008; Spielberger, 1966; Steyer et al., 1999). *Traits* are described as relatively enduring, automatic patterns of thoughts, feelings, and behaviors that are notably consistent across similar situations (McCrae & Costa, 2008; Nezlek, 2007). Constructs on the trait level range from biologically and culturally based characteristics to psychological characteristics that are conceptualized or

modeled as enduring (Nezlek, 2007). By contrast, *states* are defined as rather temporary and brief conditions that vary across time and situations (Davitz, 1969). State-level constructs can be situational variables, behaviors, and psychological states, such as mood (Nezlek, 2007). Although there is some agreement that there are psychological attributes that are more stable than others, there is no exhaustive or clear-cut taxonomy that shows which psychological constructs can be regarded as traits or states, and moreover, there is no information on how the respective constructs should be optimally operationalized (Kandler, Zimmermann, & McAdams, 2014). Moreover, it is possible to position the same construct in different places along a trait-state continuum. Probably the most prominent example is the conceptualization of anxiety as both a state and a trait (Spielberger, 1966). Although this conceptual distinction has primarily been made in reference to affective constructs, it can also be applied to other classes of psychological attributes such as personality or motivational constructs (Fleeson & Jayawickreme, 2015; Hidi & Renninger, 2006). Many theoretical approaches point to strong links between individual differences in states and individual differences in traits (Mischel, 2004). For example, in whole trait theory, traits are defined as a density distribution of states over time (Fleeson & Jayawickreme, 2015), where states have the same affective, behavioral, and cognitive content as a corresponding trait but are operationalized and assessed over shorter timespans (Zillig, Hemenover, & Dienstbier, 2002).

Measuring States and Traits

One way to operationalize state and trait characterizations of the same construct is to ask raters to base their responses on different periods of time. For instance, when an item refers to "The last 2 weeks…" versus "In general....," it points to different time frames that represent state versus trait assessments, respectively. But what exactly are the differences that emerge in response behaviors when individuals are asked about their general behavior versus their behavior over a shorter time period?

Giving an answer to an item requires a complex integration of processes, information, and attitudes. From a cognitive perspective, a different time frame can address very different underlying response processes. In survey research, the process of responding to an item includes comprehension, retrieval, judgment, and response (Tourangeau, Rips, & Rasinski, 2009). The retrieval process, in particular, can be affected by different time frames. Several characteristics of the recalled information (e.g., the length of time since an event occurred) can

affect the accuracy and completeness of the answer (Jobe, Tourangeau, & Smith, 1993). Beyond that, when an item has a longer time frame, a respondent typically uses more semantic rather than episodic knowledge to answer the item (Robinson & Clore, 2002). Besides cognitive effects based on different time frames, the environment—in terms of the number and quality of events that happened within certain time frames—also plays an important role. Whereas shorter time frames might include more intense and more specific reactions to events, they might simultaneously include fewer events in absolute numbers.

The use of different time frames is likely to go along with different analytical features. If applying trait time frames leads to higher levels of stability, and, by definition, fewer individual differences in change over time, researchers will consequently be less likely to find perturbations in traits than in states, which are supposed to be less stable. There is very little systematic research on applying and comparing different time frames across constructs. In clinical research, however, there are several inventories that include both state and trait options (Spielberger, 1966, 1995). For instance, based on Spielberger's state-trait anxiety questionnaire, a recent study observed that state anxiety revealed more state residual variance than trait anxiety and vice versa (Lance, Christie, & Williamson, 2019). In addition, the Positive and Negative Affect Schedule (PANAS) scales include explicit variations along the continuum of state and trait time frames that are directly related to the temporal stabilities and mean levels of the PANAS scales (Watson, Clark, & Tellegen, 1988). Other studies have applied different time frames to the same construct (e.g., the Big Five), for example, by testing processes between states and traits (Fleeson, 2001; Howell, Ksendzova, Nestingen, Yerahian, & Iyer, 2017). However, none of these studies have tested whether using different time frames to measure the same construct affects the interplay of two or more sets of constructs over time.

So far, no common standards have been developed regarding which time frames are appropriate for assessing either traits or states. For trait measures, this might be comparably easy to define because such measures should refer to general and typical behavior—this is why a time frame of "in general" seems to fit the definition quiet well (Robinson & Clore, 2002). By contrast, the operationalization of states is not as clear-cut. As described above, measurements of states should refer to rather brief and temporary conditions. But what is the most appropriate brief or temporary time frame that should be applied to adequately measure states? This as yet unsolved question might have more than one answer because it depends on the purpose of the study as well as the nature of the construct itself. Previous studies using self-reported state measures have differed with regard to the goal of assessing a state. Some studies

have aimed to map a history or distribution of states (e.g., Finnigan & Vazire, 2018; Neubauer, Scott, Sliwinski, & Smyth, 2019; Sowislo, Orth, & Meier, 2014). In such cases, time frames must obviously lie in the intervals of the repeated measures (e.g., daily, per hour). Other studies have addressed single measures of state constructs (e.g., Radloff, 1977; Ravens-Sieberer et al., 2001; Spielberger, 1966), such as in clinical diagnoses or in large-scale studies, where the applied time frame is less restricted by design. Over and above the empirical aim of the study, the nature of the construct itself is an important element for determining which time frame is appropriate. If a researcher is interested in current mood, it is appropriate to ask participants about the moment itself. If specific individual behavior is the main interest, it might be relevant to ask whether an individual has shown this behavior within a specific time frame, such as the last hour or the last day. Finally, other constructs require a substantial episode in order to be relevant for future feelings and behaviors, such as in clinical constructs that expect certain symptoms to occur, for example, in the last 2 weeks (American Psychiatric Association, 2013).

In the present study, we analyzed the impact of state and trait time frames in the context of self-esteem and depressive symptoms. Thus, our operationalizations of states and traits in the current study were based on the ways in which these constructs are typically operationalized. On the basis of the way in which trait self-esteem is typically assessed (Rosenberg, 1989), we chose to operationalize the trait time frame with items that were worded "In general…". On the basis of the way in which depressive symptoms are typically assessed, we chose to operationalize the state time frame by asking about "The last 2 weeks…".

Modeling States and Traits

In addition to measuring states and traits by using items with different time frames, there are statistical models that can be applied to disentangle the proportions of state and trait variance in psychological constructs. For this purpose, one prominent theory is latent state-trait (LST) theory (Steyer et al., 1999). In LST theory, any observed variable can be decomposed into a latent state variable and a measurement error component. Given multiple time points and hence multiple latent state variables, the variance of each latent state variable can be partitioned into a trait factor that represents time-invariant variance (i.e., the common variance across all time points) and a state residual factor (for each time point) that accounts for the time-point-specific variance (i.e., the variance that is not explained by the trait factor; Steyer et al., 1999). Even when constructs are originally conceptualized as traits or states and measured accordingly, studies using latent state-trait analyses have shown that all measures of psychological

constructs consist of both state residual and trait variance (Deinzer et al., 1995; Geiser, Götz, Preckel, & Freund, 2017; Rieger et al., 2017).

More recent modeling strategies have stressed the idea that disentangling proportions of state residual and trait variance is not only important when analyzing the constructs separately but also when investigating relations between constructs (Berry & Willoughby, 2017; Hamaker, Kuiper, & Grasman, 2015; Zyphur, Allison, et al., 2019). Traditionally, reciprocal relations among constructs have been investigated via cross-lagged panel models, such as in the case of self-esteem and depressive symptoms (Orth, Robins, & Roberts, 2008; Rieger, Göllner, Trautwein, & Roberts, 2016; but see Masselink et al., 2018). These models address associations that indicate whether and how between-person differences in one construct are related to between-person differences in the other construct at the next time point (controlling for previous between-person differences in the second construct). However, researchers have argued that in traditional cross-lagged panel models, it is assumed that every person varies over time around the same mean because these models account for only temporal stability but not for time-invariant differences between individuals (Hamaker et al., 2015). This would imply that individuals do not differ in their general, enduring levels of a construct across time. Alternative models explicitly model these enduring differences between individuals by combining latent state-trait models with cross-lagged panel models. Therefore, these types of models control for time-invariant individual differences (i.e., trait variance) when determining the prospective associations between constructs. Different configurations of these types of cross-lagged panel models have been proposed and are currently undergoing intensive discussions (e.g., Berry & Willoughby, 2017; Hamaker et al., 2015; Usami, Murayama, & Hamaker, 2019; Zyphur, Allison, et al., 2019; Zyphur, Voelkle, et al., 2019).

In sum, state-trait statistical models represent another way to disentangle states and traits. Considering these models in the analysis of reciprocal relations makes a crucial difference in the interpretation of the results. In the present study, our goal was to apply this modeling approach and combine it with variations in the measurement of states and traits in order to provide a systematic perspective on the individual and combined consequences of the two operationalizations of states and traits.

The Relation between Self-Esteem and Depressive Symptoms

To address the impact of the measurement and modeling of states and traits, we drew on an intensively studied research question in personality and clinical psychology (Sowislo &

Orth, 2013): What is the relation between self-esteem and depressive symptoms? In fact, the separation of states and traits seems to be very relevant for this field of research because there has been an imbalance in the measurement of the constructs as well as a focus on statistical models that do not disentangle state residual and trait variance.

The leading theoretical model for explaining the relation between self-esteem and depressive symptoms is the *vulnerability model*, which proposes that low self-esteem is a risk factor for future depressive symptoms (Beck, 1967). Accordingly, self-esteem exerts a dispositional influence on depressive symptoms when a person cannot cope with environmental strains. In contrast to the vulnerability model, the *scar model* offers a different explanation for the relation between self-esteem and depressive symptoms. The scar model proposes that low self-esteem is a consequence of depressive symptoms (Lewinsohn, Steinmetz, Larson, & Franklin, 1981). These two competing theoretical models have been studied intensively in recent decades. With respect to longitudinal and prospective designs, a large number of studies that have varied in their samples and measures have supported the vulnerability model on the interindividual level (Ormel, Oldehinkel, & Vollebergh, 2004; Orth et al., 2008; Orth, Robins, & Meier, 2009; Orth, Robins, Trzesniewski, Maes, & Schmitt, 2009; Rieger et al., 2016; Sowislo & Orth, 2013). In addition, the vulnerability pattern has been found to hold across gender, age, instruments, and the time lag between assessments (Orth et al., 2008; Rieger et al., 2016; Sowislo & Orth, 2013). Interestingly, in prior studies, self-esteem has usually revealed greater temporal stability than depressive symptoms (Orth et al., 2008; Orth, Robins, Trzesniewski et al., 2009; Rieger et al., 2016). This might be due to an imbalance in the measurement of the constructs: Self-esteem has typically been assessed as a trait ("In general…") and depressive symptoms as a state ("During the last 2 weeks…"). Therefore, one question that guided the present research was whether this imbalance has had an impact on the results.

In fact, despite the classical measures of self-esteem such as the Rosenberg Self-Esteem Scale (Rosenberg, 1989), which measures self-esteem as a trait, some authors have argued that less stable measures of self-esteem should also be included in studies on self-esteem (Brown & Marshall, 2006; Heatherton & Polivy, 1991). Along with attempts to measure self-esteem variability (Geukes et al., 2017; Kernis, 2006; Kernis, Cornell, Sun, Berry, & Harlow, 1993; Webster, Smith, Brunell, Paddock, & Nezlek, 2017), some attempts have been made to measure state self-esteem (Geukes et al., 2017; Heatherton & Polivy, 1991; Ravens-Sieberer et al., 2001). By contrast, typical measures of depressive symptoms (Radloff, 1977) represent state assessments (Spaderna, Schmukle, & Krohne, 2002; Spielberger, Ritterband, Reheiser, &

Brunner, 2003). At the same time, there have also been attempts to measure trait depressive symptoms (Spielberger, 1995). Hence, although there are examples across studies where both self-esteem and depressive symptoms were assessed with different time frames (e.g., Heatherton & Polivy, 1991; Radloff, 1977; Rosenberg, 1989; Spielberger, 1995), there have been no studies that have done so simultaneously, let alone systematic studies that have employed a longitudinal design to determine whether the different time frames affect the interplay of self-esteem and depressive symptoms.

In addition, previous research on the relation between self-esteem and depressive symptoms has mostly failed to apply models that statistically disentangle states and traits (but see Masselink et al., 2018). Yet, when analyzing self-esteem and depressive symptoms separately, both constructs have shown (varying) proportions of state residual and trait variance (Cole & Martin, 2005; Donnellan, Kenny, Trzesniewski, Lucas, & Conger, 2012; Dumenci & Windle, 1996; Wagner, Lüdtke, & Trautwein, 2016). For depressive symptoms, the proportions were already shown to vary across different instruments (Mohiyeddini, Hautzinger, & Bauer, 2002). Analyses of global self-esteem have indicated that most of the variance in self-esteem can be attributed to trait variance, whereas state residual variance explained a smaller but still substantial proportion of variance (Donnellan et al., 2012; Wagner et al., 2016). There is a recent study, which is the only one we know of, that explicitly modeled state and trait variance in the analysis of the reciprocal relations between self-esteem and depressive symptoms (Masselink et al., 2018). Across three samples, the authors showed weak vulnerability effects, and they did not observe effects that reflected the scar model. However, across all three samples, they used trait measures of self-esteem and state measures of depressive symptoms, mirroring the imbalance in state and trait measurements from studies using traditional cross-lagged panel models. An important extension would therefore be to combine both the measurement and modeling of states and traits in the reciprocal analysis of self-esteem and depressive symptoms.

The Present Research

The objective of the present research was to examine the effects of measuring and modeling states and traits in the context of an intensively studied research question addressing the reciprocal associations between self-esteem and depressive symptoms. To be more specific, we captured states and traits, first, with measures that employed different time frames to conceptualize both self-esteem and depressive symptom ratings. Second, we analytically disentangled the proportions of state residual and trait variance by using state-trait statistical models.

To our knowledge, no other study has integrated these two approaches in an attempt to disentangle states and traits let alone applied such a combined approach to analyze the relations between self-esteem and depressive symptoms. For this purpose, we designed two independent experimental, longitudinal studies that included four time points over the course of 1 semester and four conditions in a 2 × 2 design. The four conditions differed in terms of the time frame applied to self-esteem and depressive symptoms. We began our research by conducting an exploratory study, which was followed by a confirmatory replication study.

Exploratory Study

Research Questions

In the exploratory study, we addressed three overall research questions. First, given the idea that the time frame (i.e., "In general…" vs. "During the last 2 weeks…") should impact the stability of the constructs (Watson et al., 1988), we examined whether the time frame would lead to different stabilities in terms of raw rank-order stabilities as well as the division into proportions of state residual and trait variance (Research Question 1). We expected trait measures to consist of more trait variance than state measures. Second, we investigated what the reciprocal relations would be between self-esteem and depressive symptoms in the different time frame conditions (Research Question 2). We expected to replicate the results of previous studies that used the traditional approach to measure the constructs (self-esteem as a trait, depressive symptoms as a state). We did not formulate a priori hypotheses for the other conditions. Third, we analyzed the prospective relations between self-esteem and depressive symptoms by employing reciprocal models that decomposed the variance, thereby controlling for stable trait differences (Steyer & Schmitt, 1994; Zyphur, Allison, et al., 2019). We combined this approach with the differentiation of the state and trait time frames to determine the effects in the different time frame conditions (Research Question 3). This research question was exploratory as well, and therefore, we did not formulate specific hypotheses.

Method

To ensure adherence with ethical principles in the treatment of sensitive personal data, the ethics committee of the Faculty of Economics and Social Sciences of the University of Tübingen approved both studies [Reference Number: A2.5.4-068_aa].

Sample and procedure. A total of N = 648 (76% female, age M = 20.34, SD = 2.89) students at Time 1 (T1), N = 613 (78% female) students at T2, N = 600 (78% female) students

at T3, and N = 562 (77% female) students at T4 took part in the exploratory study. Data were collected between November 2016 and February 2017. All students who provided information at a minimum of one time point were included in the analyses. Students came from more than 20 different majors, with the highest percentages of students studying psychology (25.5%) or economics (14.8%); 88% of the students were born in Germany; 49% of the students' mothers and 56% of the students' fathers had a minimum of a high-school degree.

We tested whether students who participated at all time points (continuers: N = 526) differed on the study variables and other relevant outcome variables from those who participated at only the first time point (dropouts: N = 122). Results indicated no differences in the state measure of depressive symptoms, the trait measure of self-esteem, neuroticism, agreeableness, extraversion, openness, gender, and parents' education. We observed small differences (*d*s ranged from |0.30| to |0.36|) in the trait measure of depressive symptoms, the state measure of self-esteem, and high school grade point average. We observed medium differences (d = -0.60) in conscientiousness (details on the attrition analyses are presented in Supplemental Material A1). To address missing values, we used full information maximum likelihood (FIML) estimation (see also below). In addition, in order to make the missing at random (MAR) assumption more plausible, we used high school grade point average and conscientiousness as auxiliary variables in all analyses (Collins, Schafer, & Kam, 2001). To incorporate the auxiliary variables, we used the so-called "saturated correlates" model as implemented in Mplus (Asparouhov & Muthen, 2008; Graham, 2003). In this approach, auxiliary variables are not part of the structural model itself but are correlated with all variables in the model. By using this additional information (i.e., including the auxiliary variables), the precision of the estimation can be improved and can minimize biases if auxiliary variables are systematically related to the missing mechanism.

The study was conducted over the course of 1 semester. The interval between each of the four assessment was 3 weeks. The study was an online survey, and the link was sent via email. Students were mostly recruited in large, first-semester lectures from the Department of Economics and Social Sciences and the Department of Science at the University of Tübingen. In addition to recruiting in lectures, we also sent an email through the university mailing list that invited all students from the University of Tübingen to participate. Students received 25 Euro as well as feedback on their interests and personality traits in return for completing all four assessments.

In order to ensure successful randomization including balanced gender distributions in each group, we applied a conditional randomization process in two steps. First, the sample was divided by gender. Second, in both samples, students were randomly assigned to one of four conditions in a 2 x 2 between-subjects design. The time frames of the experimental conditions were varied for self-esteem and depressive symptoms (Condition A: self-esteem trait, depressive symptoms state; Condition B: self-esteem trait, depressive symptoms trait; Condition C: self-esteem state, depressive symptoms trait; Condition D: self-esteem state, depressive symptoms state). Prior to the exploratory study, we conducted a power analysis. To decide what the sample size should be and to determine the power needed to find effects in the cross-lagged panel models, we used a Monte Carlo study as implemented in Mplus (Muthén & Muthén, 2002). To be more specific, we specified a cross-lagged panel simulation model with varying sample sizes: N = 150, N = 175, and N = 200. Parameters were derived from a previous replication study on the relation between self-esteem and depressive symptoms by Rieger et al. (2016) because parameters from replication studies tend to be less biased estimates of statistical power (Anderson, Kelley, & Maxwell, 2017; Anderson & Maxwell, 2017). Moreover, the parameter estimates were very similar to the original study published by Orth et al. (2008). The population model included the following parameters: (a) temporal stabilities of self-esteem (.74, .73, .73) and depressive symptoms (.34, .31, .31) over time, (b) cross-construct associations within measurement time points (-.18, -.08, -.08, -.08), and (c) reciprocal effects between self-esteem and depressive symptoms (-.19, -.22, -.22; -.06, -.01, -.01). The results of these models indicated, respectively, a mean power level of 1.00 and .94 (N = 150), 1.00 and .97 (N = 175), and 1.00 and .98 (N = 200) for the autoregressive parameters and a mean power level of .77 (N = 150), .86 (N = 175), and .87 (N = 200) for the cross-lagged parameters from self-esteem to depressive symptoms. In the exploratory study, we recruited a total of N = 837 participants, of which N = 683 (resulting in approximately N = 170 per condition) participated in the study (see Table 1 for sample sizes per condition at each time point). Hence, the study had sufficient power to detect the expected effects.

Table 1
Group Sample Sizes

Time point	Exploratory study				Confirmatory study			
	Condition A (S trait, D state)	Condition B (S trait, D trait)	Condition C (S state, D trait)	Condition D (S state, D state)	Condition A (S trait, D state)	Condition B (S trait, D trait)	Condition C (S state, D trait)	Condition D (S state, D state)
T1	154	154	159	172	256	238	238	237
T2	138	138	157	162	250	230	237	232
T3	137	137	150	157	251	246	240	244
T4	130	130	141	147	252	235	232	223

Note. S = Self-esteem; D = Depressive symptoms; total sample sizes $N_{\text{Exploratory Study}} = 683$, $N_{\text{Confirmatory Study}} = 1{,}087$; there were students who joined the study late.

Instruments. We used data from a longitudinal, experimental, multiconstruct study that was designed to assess personality development in freshmen students. The main goal of the study was to analyze the experimental manipulation of the time frame in which self-esteem and depressive symptoms were measured. However, the study included other measures, which we present in Table A2 in Supplemental Material A2. The present investigation is the first to use this data set. In the following, we will describe in greater detail the instruments we used in the present investigation.

Depressive symptoms. Depressive symptoms were assessed with the German version ("Allgemeine Depressionsskala"; ADS; Hautzinger & Bailer, 1993) of the Center for Epidemiological Studies Depression Scale (Radloff, 1977). For this 15-item inventory, participants were asked to rate their depressive symptoms (e.g., "My sleep was restless") on a 4-point Likert scale ranging from 0 (*strongly disagree*) to 3 (*strongly agree*; due to our design, we had to ensure comparability of the item versions by changing the answer options frequencies from the original ADS)[11]. The inventory can be applied in nonclinical samples and allows a dimensional interpretation of depressive symptoms to be made (Hautzinger & Bailer, 1993). In our study, we used two different time frames for differentiating the trait and state measures of depressive symptoms. First, we applied a time frame, which asked participants to rate the statements with respect to themselves "During the last 2 weeks." For the trait assessment, we used the alternative time frame of "In general." In the trait time frame, it was necessary to make slight adaptations of two items in order to make the trait rating logical (e.g., depressive symptoms trait: "I

[11] A complete list of the original items is presented in Radloff (1977). Therefore, the items are not provided in the present manuscript. For details, see Supplemental Material A3.

talk less than in the past"; depressive symptoms state: "I talked less than usual"; we repeated the analyses without these items in the respective conditions and the results remained virtually unchanged, for details, see Supplemental Material A3). The ADS' reliability was high (McDonald's ω > .90, for details see Table 2) across all four time points for both the state and the trait measures.

Self-esteem. To measure self-esteem, we used the Rosenberg Self-Esteem Scale (Rosenberg, 1989), which has been shown to be valid for assessing global self-esteem. The scale comprises 10 statements (e.g., "I feel that I have a number of good qualities") that are self-rated on a 4-point Likert scale (1 = *strongly disagree* to 4 = *strongly agree*). Again, we used two different versions of the inventory that differed in the time frames used to measure the traits and states ("In general" vs. "During the last 2 weeks"). In this state version, it was necessary to slightly adapt one item in order to make the state rating logical (self-esteem trait: "I think I am no good at all"; self-esteem state: "I often thought I was no good"; for details, see Supplemental Material A3). Reliability was high (McDonald's ω > .90, see Table 2 for details) at all four time points for the state as well as well as for the trait measures.

Statistical Analysis

All models were estimated in the framework of multiple-group longitudinal confirmatory factor analyses and used full information maximum likelihood estimation with robust standard errors (Muthén & Muthén, 1998-2017). The significance level was set to 5% (two-tailed). For the results on the cross-lagged panel models and latent state-trait cross-lagged panel models, we report 95% confidence intervals. The dataset and all data analysis scripts are available at the Open Science Framework at the following address: https://osf.io/zu2w6/.

Parceling and model fit. In line with previous research on self-esteem and depressive symptoms (Orth et al., 2008; Rieger et al., 2016), we randomly constructed three item parcels for both self-esteem and depressive symptoms (see Supplemental Material A3 for the division of the items into parcels). According to Little, Cunningham, Shahar, and Widaman (2002), item parcels lead to more reliable latent variables than individual items (but see Marsh, Lüdtke, Nagengast, Morin, & Davier, 2013). Due to the specific variance in item parcels over time, we used the correlated-uniqueness approach (Cole, Ciesla, & Steiger, 2007). To evaluate model fit, we used the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA), and the Standardized Root Mean Square Residual (SRMR). Hu and Bentler (1999) recommended that a CFI and TLI equal to or greater than .95

and an RMSEA and SRMR equal to or less than .05 be considered indicative of a good fit (for the SRMR, values less than .09 are still acceptable). Because the models were based on the MLR estimator, comparisons of model fit were made with the Satorra-Bentler Scaled Chi-Square difference test (Satorra & Bentler, 2001). In addition, we considered differences in the goodness-of-fit indices greater than or equal to .01 to indicate model fit differences (Chen, 2007; Cheung & Rensvold, 2002).

Longitudinal measurement invariance. Before addressing our research questions, we tested for measurement invariance within the time frame conditions over time. The procedure and the results of measurement invariance testing are presented in detail in Supplemental Material B1. In sum, we assumed strong measurement invariance over time because a model with longitudinal constraints on all loadings and intercepts fit the data well (CFI = .989, TLI = .986, RMSEA = .034, SRMR = .043). Because we did not assume that the measurement properties were the same for the state and trait time frames, we did not impose measurement invariance between conditions. All models were identified by applying a "nonarbitrary" model identification process proposed by Little, Slegers, and Card (2006). Instead of using the marker-variable method by fixing a single intercept and a single factor loading to identify the models, this approach uses a weighted combination of all indicators to apply an optimal balance across indicators. By fixing the average intercept to zero and the average loading to 1, the latent variable reflects the optimally weighted metric across all indicators.

Stabilities of the measures. For Research Question 1, we investigated the stabilities of self-esteem and depressive symptoms in the state as well as the trait time frame conditions. We analyzed raw rank-order stabilities from the latent state models and applied latent state-trait models (Steyer et al., 1999) that divide the total variance of the latent state variables into a state residual part, a trait part, and an error part. Then, we compared the respective portions of variance between the state and trait measures. Before comparing the proportions of variance, we log-transformed all variances to achieve a better approximation of the normality assumption.

Cross-lagged panel model (CLPM). For Research Question 2, we analyzed the reciprocal relations between self-esteem and depressive symptoms in the different time frame conditions. We used the model most commonly used in past research to analyze the interplay of self-esteem and depressive symptoms: the (multiple-indicator, multiple-group) cross-lagged panel model (CLPM; see Figure 1). This modeling approach includes autoregressive coeffi-

cients as well as cross-lagged coefficients from one time point to another. A cross-lagged coefficient indicates whether and how differences between individuals on one variable at a specific time point predict differences between individuals in another variable at the next time point, controlling for differences between individuals at the previous time point. For example, a negative cross-lagged path from self-esteem to depressive symptoms would indicate that students with higher self-esteem than other students at a specific time point are predicted to have lower depressive symptoms at the next time point, when their previous differences in depressive symptoms are controlled for. To be able to compare different coefficients between the conditions, we focused our interpretation on standardized regression coefficients. We specified two different structural models. First, we freely estimated all the structural coefficients. Second, we constrained all autoregressive and cross-lagged coefficients to be equivalent over time. If the model fit did not decrease substantially, we preferred the more parsimonious model (a model with constraints on all autoregressive and cross-lagged coefficients). In the light of such a complex modeling procedure, the constrained model solution provides greater precision in estimating the effects because it has more degrees of freedom and allows the consistency of the findings to be tested over time (Little, Preacher, Selig, & Card, 2007; MacCallum, Browne, & Cai, 2006).

Latent state-trait cross-lagged panel model (LST-CLPM). For Research Question 3, we analyzed the reciprocal relations between self-esteem and depressive symptoms in the different time frame conditions when controlling for stable trait differences. To do so, we combined traditional cross-lagged panel models with latent state-trait models. Thus, the reciprocal relations between constructs were estimated while controlling for the trait variance in the constructs (see Figure 1). This model is a bivariate, cross-lagged version of Steyer and Schmitt's (1994) latent state-trait autoregressive model, which was also introduced by Zyphur, Allison, et al. (2019; see AR CL model with unit effects). We refer to this model as a latent state-trait cross-lagged panel model (LST-CLPM). In this model, a cross-lagged coefficient indicates whether and how differences between individuals on one variable at a specific time point predict differences between individuals in another variable at the next time point, controlling for differences between individuals at the previous time point and differences in their trait level. For example, a negative cross-lagged path from self-esteem to depressive symptoms would indicate that for students with the same trait level in depressive symptoms, those with higher self-esteem than other students at a specific time point are predicted to have

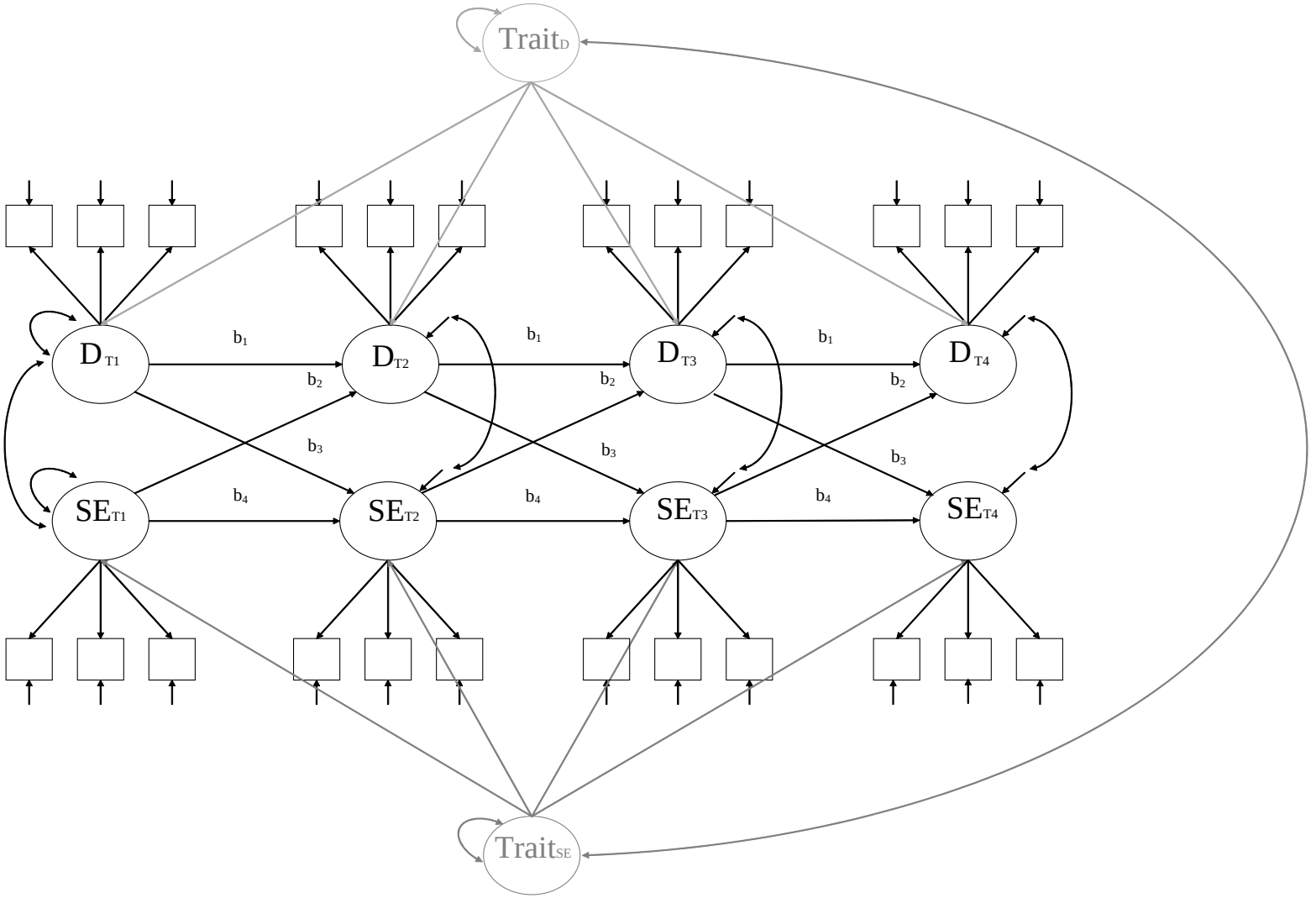


Figure 1. Cross-lagged panel model. Additional trait factor in grey. Mean structure and correlated uniquenesses are not displayed.

lower depressive symptoms at the next time point, when their previous differences in depressive symptoms are controlled for. Hence, in comparison with the autoregressive and cross-lagged regression coefficients (β) in the regular CLPM, the LST-CLPM controls for individuals' overall trait level in the variables. Therefore, we refer to these cross-lagged and autoregressive coefficients as trait-controlled regression coefficients (β_{TRC}).

Results

To analyze the impacts of the state and trait versions of self-esteem and depressive symptoms on their interplay, we began by computing descriptive statistics. Table 2 presents the means and standard deviations for the trait and state measures of self-esteem and depressive symptoms. Results indicated that, across the four time points, the means of the trait self-esteem measure ranged from M = 3.18 to 3.25 (SD = 0.54 to 0.58), and the means of the state self-esteem measure ranged from M = 3.00 to 3.05 (SD = 0.56 to 0.64). For depressive symptoms, the means of the trait measure were in the range of M = 0.90 to 0.98 (SD = 0.48 to 0.54), and the means of the state measure were in the range of M = 0.98 to 1.17 (SD = 0.52 to 0.57). Cross-sectional correlations and gender-specific descriptive results are presented in Supplemental Material B2 and B3.

Stability over time. For Research Question 1, we expected that varying the time frames of the items would be associated with the stability of the measures over the four time points. We began by inspecting the raw rank-order stabilities before evaluating the latent state-trait analyses. As shown in Table 2, the raw rank-order stabilities were higher for both of the trait measures (self-esteem: r = .92 to .96, depressive symptoms: r = .88 to .91) than for the respective state measures (self-esteem: r = .75 to .80, depressive symptoms: r = .67 to .75).

Table 2
Latent Descriptive Statistics, McDonalds ω, and *Raw Rank-Order Stabilities for all Measures*

		Exploratory study						Confirmatory study					
		Depressive symptoms			Self-esteem			Depressive symptoms			Self-esteem		
		M (SD)	ω	$r_{t,t-1}$	*M (SD)*	ω	$r_{t,t-1}$	*M (SD)*	ω	$r_{t,t-1}$	*M (SD)*	ω	$r_{t,t-1}$
T1	State	0.98 (0.52)	.90		3.05 (0.57)	.90		1.05 (0.51)	.89		2.98 (0.60)	.91	
	Trait	0.90 (0.48)	.91		3.18 (0.57)	.90		1.01 (0.52)	.91		3.05 (0.57)	.91	
T2	State	1.03 (0.56)	.90	.70	3.04 (0.56)	.92	.75	1.07 (0.52)	.90	.72	3.04 (0.55)	.91	.82
	Trait	0.93 (0.51)	.92	.88	3.18 (0.56)	.90	.93	1.03 (0.55)	.92	.88	3.06 (0.54)	.91	.93
T3	State	1.04 (0.57)	.92	.67	3.00 (0.60)	.92	.80	1.06 (0.54)	.91	.76	3.04 (0.57)	.91	.81
	Trait	0.98 (0.54)	.92	.91	3.23 (0.54)	.92	.96	1.01 (0.54)	.92	.87	3.11 (0.56)	.92	.92
T4	State	1.17 (0.57)	.93	.75	3.01 (0.64)	.91	.76	1.12 (0.54)	.91	.73	3.05 (0.59)	.92	.83
	Trait	0.93 (0.51)	.93	.89	3.25 (0.58)	.91	.92	1.00 (0.56)	.93	.86	3.13 (0.57)	.92	.92

Note. Exploratory study*: N* = 683; Confirmatory study: *N* = 1,087.

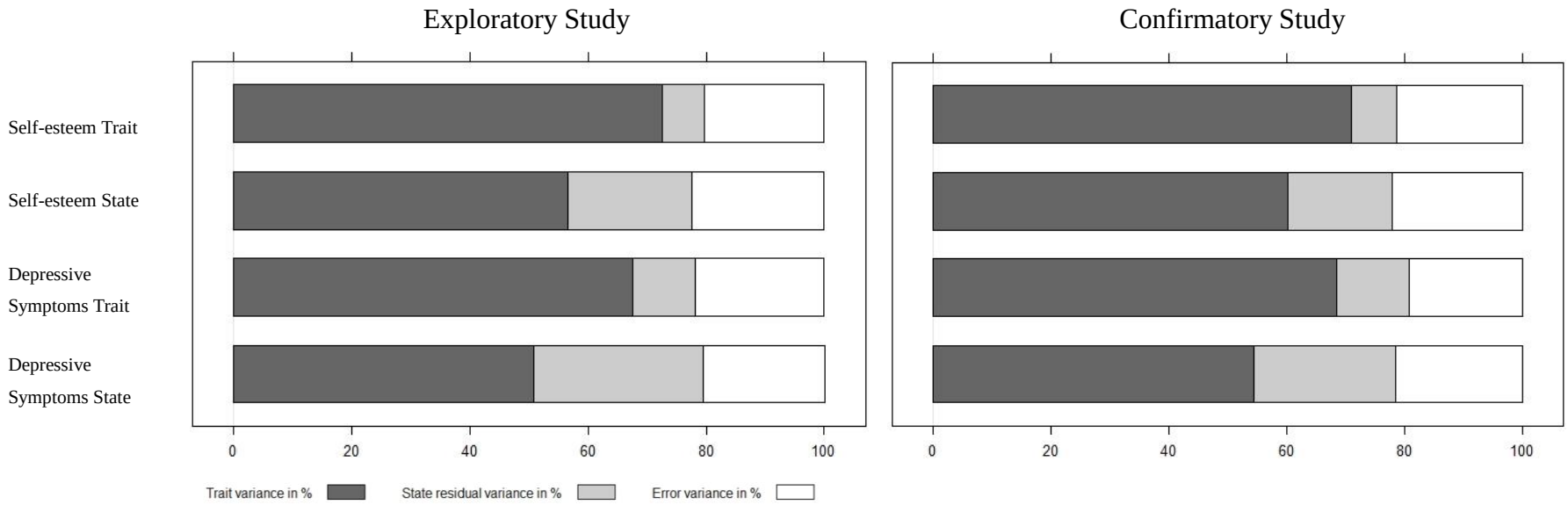


Figure 2. Decomposition of variance for the state and the trait measures of both self-esteem and depressive symptoms in both studies (in percentages). The exact proportions of variance for the trait variance/state residual variance/error variance for self-esteem trait: 72.6/7.1/20.3% (exploratory study) and 71.0/7.7/21.3% (confirmatory study); Self-esteem state: 56.5/21.1/22.3% (exploratory study) and 60.2/17.7/22.1% (confirmatory study); Depressive symptoms trait: 67.5/10.7/21.8% (exploratory study) and 68.5/12.2/19.3% (confirmatory study); Depressive symptoms state: 50.8/28.7/20.5% (exploratory study) and 54.4/24.0/21.6% (confirmatory study).

Next, we performed latent state-trait analyses, which corroborated the findings from the rank-order correlations. Figure 2 depicts the proportions of variance (as percentages). For depressive symptoms, the trait measure consisted of 67.5% trait variance, 10.7% state residual variance, and 21.8% error variance. By contrast, the state measure consisted of 50.8% trait variance, 28.7% state residual variance, and a comparable percentage of error variance (20.5%). The state and trait measures differed significantly in amount of trait (Δ = 16.7%, logΔ = 0.284, *SE* = 0.063, $p < .001$) and state (Δ = -18.0%, logΔ = -0.986, *SE* = 0.126, $p < .001$) variance. For self-esteem, the trait measure contained 72.6% trait variance, 7.1% state residual variance, and 20.3% error variance, whereas the state measure contained 56.5% trait variance, 21.1% state residual variance, and 22.3% [12] error variance. Again, the state and trait measures differed significantly in amount of trait (Δ = 16.1%, logΔ = 0.250, *SE* = 0.054, $p < .001$) and state (Δ = -14.0%, logΔ = -1.091*;* *SE* = 0.152, $p < .001$) variance. Hence, in line with our expectations, both trait measures showed higher proportions of trait variance and could consequently be considered to have more stability over time.

Reciprocal relations between self-esteem and depressive symptoms. With Research Question 2, we aimed to analyze the prospective relations between self-esteem and depressive symptoms in the four time frame conditions. We estimated cross-lagged panel regression models and compared the standardized structural coefficients between the four conditions. All models fit the data well (see Table B1 in Supplemental Material B1). We focused our results on a model with longitudinal constraints on all autoregressive and cross-lagged coefficients because the additional restrictions did not lead to a significantly worse fit. The results for a model in which all structural coefficients were freely estimated are reported in Table B9 in Supplemental Material B4.

Table 3 presents all standardized regression coefficients and all confidence intervals for the cross-lagged panel models from the four conditions. In the following, we describe the standardized regression coefficients from the cross-lagged panel models that were statistically significant. In all four conditions, all autoregressive coefficients were statistically significant. The autoregressive coefficients in the state measures were lower than the ones in the trait measures for both self-esteem (state measures: β = .53 - .59, trait measures: β = .90 - .99) and depressive symptoms (state measures: β = .47 - .58, trait measures: β = .76 - .81).

[12] Due to rounding in Mplus, these percentages do not add up to 100%.

Table 3

Estimates [and 95% Confidence Intervals] of the Structural Coefficients from the Cross-Lagged Panel Model (CLPM)

		Exploratory study				Confirmatory study			
		A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state	A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state
Coefficient		β [CI]	β [CI]	β [CI]	β [CI]	β [CI]	β [CI]	β [CI]	β [CI]
Stability									
Self-esteem	T1→T2	**.96 [.89, 1.03]**	**.96 [.89, 1.04]**	**.59 [.28, .90]**	**.55 [.36, .78]**	**1.01 [.96, 1.06]**	**.99 [.93, 1.05]**	**.65 [.49, .82]**	**.86 [.74, .99]**
	T2→T3	**.98 [.92, 1.04]**	**.99 [.92, 1.07]**	**.56 [.23, .89]**	**.57 [.35, .80]**	**1.00 [.96, 1.05]**	**.97 [.91, 1.03]**	**.60 [.44, .77]**	**.86 [.73, .99]**
	T3→T4	**.95 [.88, 1.02]**	**.90 [.82, .99]**	**.55 [.24, .86]**	**.53 [.32, .74]**	**.99 [.94, 1.04]**	**.97 [.91, 1.04]**	**.59 [.42, .76]**	**.85 [.72, .98]**
Depressive symptoms	T1→T2	**.47 [.29, .64]**	**.77 [.65, .89]**	**.77 [.60, .94]**	**.54 [.35, .73]**	**.56 [.44, .68]**	**.78 [.70, .87]**	**.73 [.58, .88]**	**.66 [.51, .80]**
	T2→T3	**.47 [.31, .64]**	**.77 [.65, .90]**	**.81 [.64, .97]**	**.54 [.35, .74]**	**.55 [.42, .68]**	**.81 [.74, .89]**	**.69 [.54, .85]**	**.68 [.51, .84]**
	T3→T4	**.50 [.32, .68]**	**.76 [.65, .88]**	**.81 [.64, .97]**	**.58 [.38, .78]**	**.56 [.44, .69]**	**.79 [.71, .88]**	**.70 [.58, .83]**	**.66 [.50, .81]**
Cross-lagged									
	SE1→D2	**-.33 [-.49, -.17]**	**-.16 [-.28, -.04]**	-.12 [-.29, .05]	-.16 [-.36, .04]	**-.23 [-.34, -.12]**	**-.12 [-.20, -.04]**	**-.20 [-.35, -.06]**	-.12 [-.27, .03]
	SE2→D3	**-.32 [-.49, -.16]**	**-.16 [-.28, -.04]**	-.12 [-.28, .04]	-.15 [-.34, .04]	**-.23 [-.33, -.12]**	**-.12 [-.19, -.04]**	**-.18 [-.30, -.05]**	-.12 [-.27, .03]
	SE3→D4	**-.33 [-.49, -.16]**	**-.15 [-.26, -.03]**	-.12 [-.28, .05]	-.15 [-.35, .04]	**-.23 [-.34, -.12]**	**-.11 [-.19, -.04]**	**-.17 [-.29, -.04]**	-.12 [-.26, .03]
	D1→SE2	.02 [-.05, .09]	.02 [-.06, .09]	*-.29* [a] *[-.61, .03]*	**-.21 [-.39, -.02]**	**.08 [.02, .14]**	*.05 [-.01, .11]*	**-.24 [-.40, -.07]**	.02 [-.10, .14]
	D2→SE3	.02 [-.05, .10]	.02 [-.06, .10]	*-.30* [a] *[-.62, .03]*	**-.22 [-.42, -.02]**	**.08 [.02, .17]**	*.05 [-.01, .12]*	**-.24 [-.40, -.07]**	.02 [-.10, .15]
	D3→SE4	.02 [-.05, .10]	.02 [-.06, .09]	*-.29* [a] *[-.60, .03]*	**-.22 [-.42, -.02]**	**.08 [.02, .14]**	*.05 [-.01, .12]*	**-.25 [-.41, -.08]**	.02 [-.10, .15]

Note. Exploratory study: N_{Total} = 683; N_A = 179; N_B = 163; N_C = 170; N_D = 171. Confirmatory study: N_{Total} = 1,087; N_A = 280; N_B = 269; N_C = 273; N_D = 265. Condition A: Self-esteem trait, Depressive symptoms state; Condition B: Self-esteem trait, Depressive symptoms trait; Condition C: Self-esteem state, Depressive symptoms trait; Condition D: Self-esteem state, Depressive symptoms state. Bold indicates that the 95% confidence interval did not include zero. Italics indicate that the 90% confidence interval did not include zero. All coefficients are standardized.

[a] The regression coefficients in Condition C were larger than in Condition D. Also the standard errors and the respective confidence intervals were larger in Condition C than in Condition D. This is because the calculation of the standard errors depended on multiple components such as the amount of variance explained, sample size, number of predictors, and the standard deviations of the outcome and predictor (for the formula, see Cohen, Cohen, West, & Aiken, 2003)

Table 4

Estimates [and 95% Confidence Intervals] of the Structural Coefficients from the Latent State-Trait Cross-Lagged Panel Model (LST-CLPM)

		Exploratory study				Confirmatory study			
		A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state	A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state
Coefficient		β_{TRC} [CI]	β_{TRC} [CI]	β_{TRC} [CI]	β_{TRC} [CI]	β_{TRC}[CI]	β_{TRC}[CI]	β_{TRC}[CI]	β_{TRC}[CI]
Stability									
Self-esteem	T1→T2	.00 [-.15, .15]	-.05 [-.20, .11]	.21 [-.24, .65]	-.26 [-.55, .03]	-.03 [-.14, .08]	.06 [-.12, .23]	**.32 [.03, .61]**	.01 [-.25, .26]
	T2→T3	.00 [-.15, .16]	-.05 [-.20, .10]	.20 [-.23, .63]	-.31 [-.62, .01]	-.03 [-.13, .08]	.06 [-.12, .23]	**.32 [.03, .61]**	.01 [-.23, .25]
	T3→T4	.00 [-.14, .15]	-.04 [-.18, .10]	.20 [-.23, .63]	-.25 [-.50, .00]	-.03 [-.14, .08]	.06 [-.12, .23]	**.29 [.01, .57]**	.01 [-.24, .26]
Depressive symptoms	T1→T2	.34 [-.15, .82]	.11 [-.23, .44]	.03 [-.20, .25]	**.34 [.06, .62]**	**.31 [.14, .49]**	.12 [-.05, .28]	-.11 [-.27, .05]	.11 [-.14, .35]
	T2→T3	.36 [-.09, .80]	.11 [-.23, .44]	.03 [-.22, .28]	**.40 [.10, .70]**	**.28 [.11, .45]**	.12 [-.06, .30]	-.12 [-.29, .06]	.11 [-.13, .35]
	T3→T4	.36 [-.15, .87]	.11 [-.23, .44]	.03 [-.22, .28]	**.38 [.07, .68]**	**.30 [.13, .46]**	.12 [-.05, .29]	-.11 [-.28, .06]	.11 [-.14, .36]
Cross-lagged									
	SE1→D2	.13 [-.28, .53]	-.01 [-.32, .31]	-.17 [-.43, .10]	.20 [-.07, .47]	**.22 [.03, .40]**	-.00 [-.16, .16]	**-.20 [-.38, -.02]**	-.10 [-.34, .15]
	SE2→D3	.13 [-.26, .51]	-.01 [-.31, .30]	-.17 [-.44, .11]	.22 [-.08, .53]	**.20 [.03, .36]**	-.00 [-.16, .16]	**-.19 [-.37, -.01]**	-.09 [-.32, .14]
	SE3→D4	.12 [-.26, .51]	-.01 [-.30, .28]	-.17 [-.46, .12]	.20 [-.07, .47]	**.20 [.04, .37]**	-.00 [-.16, .15]	**-.18 [-.35, -.00]**	-.10 [-.34, .15]
	D1→SE2	-.08 [-.21, .06]	-.07 [-.22, .09]	.03 [-.31, .38]	**-.37 [-.60, -.14]**	-.03 [-.11, .05]	.02 [-.13, .19]	**.24 [.03, .46]**	-.14 [-.38, .09]
	D2→SE3	-.08 [-.22, .06]	-.07 [-.23, .09]	.04 [-.33, .40]	**-.46 [-.74, -.18]**	-.03 [-.10, .04]	.03 [-.14, .19]	**.27 [.04, .50]**	-.14 [-.37, .09]
	D3→SE4	-.08 [-.22, .06]	-.06 [-.22, .09]	.03 [-.31, .38]	**-.39 [-.64, -.14]**	-.03 [-.11, .05]	.03 [-.13, .19]	**.26 [.04, .48]**	-.15 [-.39, .10]

Note. Exploratory study: N_{Total} = 683; N_A = 179; N_B = 163; N_C = 170; N_D = 171. Confirmatory study: N_{Total} = 1,087; N_A = 280; N_B = 269; N_C = 273; N_D = 265. Condition A: Self-esteem trait, Depressive symptoms state; Condition B: Self-esteem trait, Depressive symptoms trait; Condition C:

Self-esteem state, Depressive symptoms trait; Condition D: Self-esteem state, Depressive symptoms state. Bold indicates that the 95% confidence interval did not include zero. Italics indicate that the 90% confidence interval did not include zero. All coefficients are standardized.

Regarding the cross-lagged coefficients, first, we aimed to replicate previous results on the prospective relations between self-esteem and depressive symptoms with the traditional measures (Condition A: self-esteem trait, depressive symptoms state). In this condition, all cross-lagged paths from self-esteem to depressive symptoms were statistically significant (β = -.33, -.32, -.33), but the paths from depressive symptoms to self-esteem were not. Similarly, in Condition B (self-esteem trait, depressive symptoms trait), all cross-lagged paths from self-esteem to depressive symptoms were statistically significant (β = -.16, -.16, -.15, but the paths were smaller than in Condition A). In Condition B, the paths from depressive symptoms to self-esteem were not statistically significant. In the two conditions in which self-esteem was assessed as a state (Conditions C and D), the cross-lagged paths from self-esteem to depressive symptoms were not statistically significant. In Condition D (self-esteem state, depressive symptoms state), the paths from depressive symptoms to self-esteem were statistically significant (Condition D: β = -.21, -.22, -.22). In Condition C (self-esteem state, depressive symptoms trait), a similar pattern occurred, but the cross-lagged paths from depressive symptoms to self-esteem were not statistically significant.

Trait-controlled reciprocal relations between self-esteem and depressive symptoms. With Research Question 3, we aimed to analyze the reciprocal relations between self-esteem and depressive symptoms when controlling for stable trait differences. To do so, we combined latent state-trait models with cross-lagged panel models. We combined this approach with the measurement of states and traits and were thus able to compare the results between the time frame conditions. Paralleling the procedure from the regular cross-lagged model, we imposed longitudinal constraints on all autoregressive and cross-lagged coefficients. The results for the model in which all structural coefficients were freely estimated are reported in Table B10 in Supplemental Material B4. The models fit the data well (see Table B1 in Supplemental Material B1). The trait factors for self-esteem and depressive symptoms were strongly correlated in all conditions (Condition A: r = -.91, Condition B: r = -.86, Condition C: r = -.93, Condition D: r = -.91). Table 4 presents the trait-controlled standardized regression (β_{TRC}) coefficients and confidence intervals for the four conditions. In Conditions A, B, and C, neither the autoregressive coefficients nor the cross-lagged paths were statistically significant. In Condition D (self-esteem state, depressive symptoms state), the autoregressive coefficients for depressive symptoms (β_{TRC} = .34, .40, .38) and the cross-lagged paths from depressive symptoms to self-esteem were statistically significant (β_{TRC} = -.37, -.46, -.39), but the paths from self-

esteem to depressive symptoms were not. In sum, when analyzing cross-lagged effects by controlling for stable trait differences, we revealed substantial cross-lagged and autoregressive paths only when both constructs were measured as states.

Summary

In sum, the exploratory study showed that the stability of the constructs and the pattern of results for the relation between self-esteem and depressive symptoms strongly depends on the time frame in which the constructs are measured and the statistical model that is applied. In the CLPM, we found support for the vulnerability model when self-esteem was measured with a trait time frame. However, a new picture emerged when we used state forms of both constructs, a picture that is similar to predictions made by the scar model. When modeling cross-lagged effects while controlling for stable trait differences using the LST-CLPM, the vulnerability effect was not supported in any of the conditions. However, in this model, we found significant paths from depressive symptoms to self-esteem when using state measures.

Confirmatory Study

Following the exploratory study, we performed a confirmatory direct replication study where we aimed to address three overarching goals: (a) to reduce the potential of false positive results from the exploratory study (Simmons, Nelson, & Simonsohn, 2011), (b) to increase the power to obtain the effects by increasing the sample size (Anderson & Maxwell, 2017), and (c) to provide a confirmatory research framework by preregistering hypotheses that were based on the exploratory study and explicitly testing them.

Hypotheses

Before we conducted the study, we preregistered the research design, procedure, statistical analyses, and hypotheses in the Open Science Framework (https://osf.io/4293g). The study was designed as a direct replication study that paralleled all possible steps from the exploratory study. On the basis of the results of the exploratory study, we formulated hypotheses for each of the previous research questions:

Research Question 1 (Stability over Time):

- Hypothesis 1a: In line with the findings from the exploratory study, we expected the trait measure of self-esteem to consist of more trait variance and thereby to exhibit higher raw rank-order stabilities than the state measure of self-esteem.

- Hypothesis 1b: In line with the findings from the exploratory study, we expected the trait measure of depressive symptoms to consist of more trait variance and thereby to exhibit higher raw rank-order stabilities than the state measure of depressive symptoms.

Research Question 2 (Reciprocal Relations):

- Hypothesis 2a: On the basis of previous research (Orth et al., 2008; Sowislo & Orth, 2013) and in line with the findings from the exploratory study, we expected that self-esteem would negatively predict depressive symptoms when self-esteem was measured as a trait. This effect was expected to be higher than the effect of depressive symptoms on trait self-esteem. This hypothesis applies to Condition A (self-esteem trait, depressive symptoms state) and Condition B (self-esteem trait, depressive symptoms trait).
- Hypothesis 2b: In line with the findings from the exploratory study, we expected that depressive symptoms would negatively predict self-esteem when self-esteem was measured as a state. This effect was expected to be higher than the effect of state self-esteem on depressive symptoms. This hypothesis applies to Condition C (self-esteem state, depressive symptoms trait) and Condition D (self-esteem state, depressive symptoms state).

Research Question 3 (Trait-Controlled Reciprocal Relations):

- Hypothesis 3a: In line with the findings from the exploratory study, we expected longitudinal predictions between self-esteem and depressive symptoms only when both constructs were measured as states (Condition D: Self-esteem state, depressive symptoms state).
- Hypothesis 3b: In line with the findings from the exploratory study, we expected that depressive symptoms would negatively predict self-esteem when both constructs were measured as states. This effect was expected to be higher than the effect of state self-esteem on state depressive symptoms. This hypothesis applies to Condition D (Self-esteem state, depressive symptoms state).

Method

The study design approved for the exploratory study [Reference Number: A2.5.4-068_aa] was used again for the confirmatory study, for which we received a renewed agreement from the ethics committee of the Faculty of Economics and Social Sciences of the University

of Tübingen. The confirmatory study paralleled all steps of the exploratory study with minor deviations in the procedure and composition of the sample that are described in detail in the following sections (for an overview of the differences between the two studies, see Supplemental Material A4).

Sample and procedure. A total of N = 1,009 (72% female, age M = 21.52, SD = 3.34) students at T1, N = 949 (73% female) students at T2, N = 947 (74% female) students at T3, and N = 936 (72% female) students at T4 took part in the confirmatory study. The data were collected from May 2019 to July 2019. All students who provided information at a minimum of one time point were included in the analyses. Students came from more than 30 different majors, with the highest percentages of students studying medicine (12.3 %), law (12.0 %), and education (10.7 %); 87% of the students were born in Germany; 50% of the students' mothers and 53% of the students' fathers had a minimum of a high-school degree.

Again, we tested whether students who completed all time points (continuers: N = 792) differed on the study variables and other relevant outcome variables from those who completed only the first time point (N = 217). Results indicated no differences in neuroticism, extraversion, agreeableness, openness, parents' education, the state and the trait measures of depressive symptoms, or the state and the trait measures of self-esteem. However, we observed small differences in high school grade point average (d = 0.15), gender (OR = 0.666, Pseudo R^2 = 0.007, see McKelvey & Zavoina, 1975), and conscientiousness (d = -0.33). Details on the attrition analysis are presented in Table A1 in Supplemental Material A1. Paralleling the procedure used in the exploratory study, we used full information maximum likelihood (FIML) estimation (see below) and added high school grade point average, conscientiousness, and gender as auxiliary variables[13] in all analyses (for details on this approach, see the exploratory study).

The design of the study was directly parallel to the design of the exploratory study, with only small deviations (for details, see Supplemental Material A4). Most importantly, in the confirmatory study, students were mostly recruited in large, second-semester lectures from all departments at the University of Tübingen. We recruited from the same majors as in the exploratory study (e.g., psychology, economics, education, media science) but additionally in-

[13] In the preregistration, we stated that we would use conscientiousness and grade point average as auxiliary variables. However, in the confirmatory study, gender also differed substantially between dropouts and continuers, and this is why we added gender as an auxiliary variable. However, the results were virtually unchanged in a model with only the previously stated auxiliary variables.

cluded other majors (e.g., law, medicine, pharmacy, rhetoric) to increase the sample size. However, this led to a slightly different sample composition in comparison with the exploratory study (for details, see the descriptions of the samples). Again, we also sent an email through the university mailing list that invited all students from University of Tübingen to participate. In the confirmatory study, in return for completing all four assessments, students received individual feedback on their personality profile, and, they were given the opportunity to take part in a lottery where a total of three iPads and 30x30 Euro gift cards were raffled.

In line with the exploratory study, we conducted a conditional randomization process, but we extended it by adding one more step. Along with balancing gender, in the confirmatory study, we also balanced students' major because we systematically recruited students from a larger variety of majors with the goal of achieving greater heterogeneity in the sample. We used the same conditions for the 2 × 2 between-subjects design as in the exploratory study. In the confirmatory study, we recruited a total of N = 1,405 participants in order to achieve at least a similar and preferably a larger number of participants in comparison with the exploratory study (see, e.g., Anderson, Kelly, & Maxwell, 2017). In the end, N = 1,087 students participated in the confirmatory study. Thus, we ended up with approximately N ~ 250 students per condition at T1 (see Table 1 for the sample sizes per condition at each time point).

Instruments. The data used in the confirmatory study stemmed from a longitudinal, experimental multiconstruct study that was designed as a direct replication of the exploratory study. The present investigation was the first to use this data set. Self-esteem and depressive symptoms were assessed with the same instruments and time frames used in the exploratory study (for details, see the exploratory study and Supplemental Material A3). Reliabilities were high for both self-esteem (McDonald's ω > .91) and depressive symptoms (McDonald's ω > .89, for details, see Table 2) at all time points and for both the state and trait versions of the measures. The data set included additional measures that were not part of the present investigation. These are presented in Supplemental Material A2.

Statistical analysis. We paralleled all statistical approaches from the exploratory study, such as the same division into item parcels (see Supplemental Material A3), the nonarbitrary model identification (Little et al., 2006), the correlated-uniqueness approach (Cole et al., 2007), the criteria for evaluating model fit, and the difference tests (for details, see the exploratory study). Again, before addressing our hypotheses, we tested for measurement invariance within the time frame conditions over time. The procedure and the results of measurement invariance

testing are presented in detail in Supplemental Material B1. In sum, we assumed strong measurement invariance over time because, a model with longitudinal constraints on all loadings and intercepts fit the data well (CFI = .990, TLI = .987, RMSEA = .033, SRMR = .035).

To test our hypotheses, we exactly paralleled all statistical analyses from the exploratory study, including rank-order stabilities and latent state-trait analyses (Hypothesis 1a, 1b), multiple-group cross-lagged panel models (Hypothesis 2a, 2b), and multiple-group latent state-trait cross-lagged panel models (Hypothesis 3a, 3b)[14]. Again, the dataset and all data analysis scripts are available at the Open Science Framework at the following address: https://osf.io/zu2w6/.

Results

In line with the procedure used in the exploratory study, we began by analyzing descriptive statistics. Table 2 presents the means and standard deviations for the trait and state measures of self-esteem and depressive symptoms for the confirmatory study. In accordance with the exploratory study, descriptive results indicated that, across the four time points, the means of the trait self-esteem measure were higher than the means of the state self-esteem measure, and the means of the trait depressive symptoms measure were lower than the means of the state depressive symptoms measure (for cross-sectional correlations and descriptive statistics separated by gender, see Supplemental Material B2 and B3).

Stability over time. Hypotheses 1a and 1b proposed that measures with trait time frames would be associated with higher stability across the four time points than the measures with state time frames. As shown in Table 2 and Figure 2, the raw rank-order stabilities as well as the variance decomposition based on the latent state-trait analyses were in line with Hypotheses 1a and 1b. Thus, the raw rank-order stabilities were higher for both of the trait measures than for the respective state measures. Correspondingly, the trait measures showed higher proportions of trait variance than the state measures (Depressive symptoms: $\Delta = 14.1\%$, $\log\Delta = 0.232$, $SE = 0.047$, $p < .001$; Self-esteem: $\Delta = 10.8\%$, $\log\Delta = 0.165$, $SE = 0.041$, $p < .001$), and the state measures showed higher proportions of state residual variance than the trait measures

[14] In the preregistration, we stated that we would run random intercept cross-lagged panel models. However, we had preregistered the syntax files for running models where the cross-lagged and autoregressive effects were estimated on the basis of the latent state variables while controlling for the trait differences (latent state-trait cross-lagged panel models). We now want to resolve this inconsistency by emphasizing that we are running the latter. In the Supplemental Material B6, we described in detail the similarities and differences between the two models. In addition, for reasons of transparency and completeness, we added the results of the random intercept cross-lagged panel models.

(Depressive symptoms: Δ = -11.8%, logΔ = -0.680, *SE* = 0.114, $p < .001$; Self-esteem: Δ = -10.0%, logΔ = -0.835*; SE* = 0.131, $p < .001$).

Reciprocal relations between self-esteem and depressive symptoms. To test Hypotheses 2a and 2b, we analyzed the reciprocal relations between self-esteem and depressive symptoms in the four time frame conditions. All cross-lagged panel models fit the data well (see Table B2 in Supplemental Material B1), and we focused on a model with longitudinal constraints on all autoregressive and cross-lagged coefficients. Results for the model in which all structural coefficients were freely estimated are reported in Table B9 in Supplemental Material B4. Table 3 displays all standardized regression coefficients and all confidence intervals for the cross-lagged panel models from the four conditions. In line with the exploratory study, for both self-esteem and depressive symptoms, the autoregressive coefficients based on the state measures were lower than those based on the trait measures. The cross-lagged coefficients in the four conditions revealed a pattern that was largely similar to the one from the exploratory study. Thus, in the two conditions where self-esteem was measured with a trait time frame (Conditions A and B), the standardized regression coefficients from self-esteem to depressive symptoms were higher than the coefficients going in the opposite direction, supporting Hypothesis 2a. When self-esteem was measured with a state time frame and depressive symptoms with a trait time frame (Condition C), the paths from depressive symptoms to self-esteem were significant and higher than the opposite paths. However, this was not the case when both constructs were measured with a state time frame (Condition D), and therefore the results only partly confirmed Hypothesis 2b.

Trait-controlled reciprocal relations between self-esteem and depressive symptoms. With Hypotheses 3a and 3b, we aimed to test the reciprocal relations between self-esteem and depressive symptoms in the different time frame conditions when controlling for stable trait differences. All latent state-trait cross-lagged panel models fit the data well (see Table B2 in Supplemental Material B1). Therefore, we imposed longitudinal constraints on all autoregressive and cross-lagged coefficients. The results for the model in which all structural coefficients were freely estimated are reported in Table B10 in Supplemental Material B4[15]. In line with the exploratory study, the trait factors for self-esteem and depressive symptoms were

[15] Results based on models without longitudinal constraints on all regression coefficients deviated slightly from those with longitudinal constraints (see Table B9 and Table B10 in Supplemental Material B4). In addition, we added the (latent state-trait) cross-lagged results for females only (the constrained and unconstrained CLPM for the exploratory and confirmatory studies and LST-CLPM for the confirmatory study; see Supplemental Material B5, Tables B11, B12, and B13). The results were virtually unchanged in comparison with the total sample.

strongly correlated in all conditions (Condition A: r = -.86, Condition B: r = -.86, Condition C: r = -.94, Condition D: r = -.84).

Table 4 presents the trait-controlled standardized regression coefficients and confidence intervals for the latent state-trait cross-lagged panel models of the four conditions in the confirmatory study. In general, most of the autoregressive and cross-lagged coefficients from the confirmatory study fell within the 95% confidence intervals from the exploratory study and vice versa (see Table 4). However, the results differed substantially with regard to reaching the *p*-value threshold. Thus, in the confirmatory study, the autoregressive coefficients were significant for the depressive symptoms state measure in Condition A and the self-esteem state measure in Condition C. The cross-lagged coefficients from self-esteem to depressive symptoms were significant in Conditions A and C, with positive coefficients in Condition A and negative coefficients in Condition C. In addition, in Condition C, depressive symptoms positively predicted self-esteem. All other cross-lagged coefficients were not statistically significant. Therefore, we did not find support for either Hypothesis 3a or Hypothesis 3b.

In sum, the confirmatory study again showed that the measuring and modeling of states and traits is substantially related to the longitudinal interplay between self-esteem and depressive symptoms. The results regarding the stabilities (Hypotheses 1a, 1b) and the reciprocal relations in the four time frame conditions (Hypotheses 2a, 2b) were largely replicated, whereas the results regarding the trait-controlled reciprocal relations (Hypotheses 3a, 3b) were comparable to but less consistent with those from the exploratory study.

General Discussion

In the present research, we examined the influence of measuring and modeling states and traits in the context of an intensively studied research question, namely, the longitudinal relation between self-esteem and depressive symptoms. For this purpose, we used longitudinal experimental designs in an exploratory study and in a confirmatory replication study. In both studies, we randomly assigned participants to one of four experimental conditions that differed with respect to the time frames (i.e., trait: “In general…” vs. state: “During the last 2 weeks…”) used to assess self-esteem and depressive symptoms. We combined this approach with statistical models that statistically decomposed the state residual and trait variance. The procedure we used was as follows: In the exploratory study, as this was the first study that ever used both

state and trait versions of both self-esteem and depressive symptoms, we had only a few expectations about the results. In the confirmatory study, we preregistered hypotheses that were based on the results of the exploratory study, and we explicitly tested these.

In general, the results from the exploratory and confirmatory studies were largely similar regarding the overall pattern of results, but some differences were identified. Across the two studies, first, we showed that using trait time frames for both self-esteem and depressive symptoms revealed higher stabilities and consequently higher amounts of trait variance than state measures of the respective constructs (Research Question 1). Second, in both studies, the reciprocal relation between self-esteem and depressive symptoms mostly depended on the time frame used to measure self-esteem: When self-esteem was measured as a trait, it predicted depressive symptoms, independent of the time frame used to measure depressive symptoms. However, when self-esteem was measured as a state, there were indications that depressive symptoms predicted state self-esteem (Research Question 2). Finally, when controlling for stable trait differences, relations between self-esteem and depressive symptoms indicated differences between the time frame conditions. Even though the overall pattern of results was comparable across the two studies, the results were less clear than the results from the first two research questions (Research Question 3).

Different Time Frames, Different Results

By explicitly presenting certain time frames in questionnaires, we can approximate a measurement-based operationalization of states and traits. Different time frames can stimulate different cognitive processes and might be related to different analytical features (e.g., stabilities); however, the impact of such a measurement decision has never been studied systematically, especially not when exploring the longitudinal relation between constructs. The present research filled this gap and showed that experimentally manipulating the measurement of states versus traits had profound consequences. Using self-esteem and depressive symptoms as example constructs to which we applied different time frames, we observed higher stabilities when applying trait measures compared with state measures. These finding are in line with results from previous studies (Lance et al., 2019; Watson et al., 1988) for which more trait variance and higher stabilities were found when using more trait-like time frames. In addition, we observed differences in reciprocal relations between self-esteem and depressive symptoms when state and trait measures of both constructs were varied. Not only was the experimental

manipulation of measuring states and traits relevant in terms of the magnitudes of the coefficients, but it was even related to the directions of the cross-lagged paths.

The way the constructs were measured is, in part, a reflection of how the constructs are conceptualized and theoretically understood. It is appropriate to ask whether it is legitimate to consider constructs (e.g., self-esteem and depressive symptoms) as both traits and states. In our example, past theory would largely argue that self-esteem is a trait (Rosenberg, 1979, 1989), and depressive symptoms reflect a state (Radloff, 1977). However, there have been deviations from these strict conceptualizations, which have emphasized the relevance of temporary deviations in self-esteem (Heatherton & Polivy, 1991) and a general disposition to exhibit depressive symptoms (Spielberger, 1995). In line with previous studies (Cole & Martin, 2005; Donnellan et al., 2012; Dumenci & Windle, 1996; Wagner et al., 2016), the present findings on latent state-trait analyses call into question the clear classification of self-esteem and depressive symptoms as states or traits only: Our findings indicate that trait measures contain state residual variance, and state measures contain trait variance. When self-esteem and depressive symptoms were assessed as states, over 50% of the variance could be attributed to trait variance. This amount of trait variance should be a warning to researchers who expect that using state time frames for items will automatically result in a preponderance of state residual variance in a measure.

In the present study, we used a conceptualization of state measures that referred to the last 2 weeks. This time frame was chosen because it mirrored the state conceptualization of depressive symptoms, which was a starting point for our research. However, we want to emphasize that this is just one possible way to conceptualize a state construct—and also generally speaking, a conservative one. Given the nature of a construct allows certain conceptualizations of states, it will be necessary to extend the current research to determine the consequences of more narrow time frames (e.g., referring to the last day, the last moment, an ambulatory assessment of actual behavior, or physiological measures). When applying the same intervals of time between measurement occasions, we would expect that the shorter the time frames used in a question, the less stable the measures of the constructs would be (Watson et al., 1988). However, the extent to which certain differences in time frames contribute to construct stability is not yet well understood, raising the question of whether increasing the intervals between measurement occasions would lead to a linear increase in the stabilities (i.e., proportions of state residual and trait variance). Due to the fact that the present results on the relation between self-esteem and depressive symptoms mostly depended on the time frame used to assess self-

esteem, it would be particularly insightful to figure out whether and how the pattern of results would change when other variations in the measurement of self-esteem were applied. If we apply the conclusions of the present study, we would expect that shorter time frames in measuring state self-esteem (e.g., "the last day" or "the last hour") would make it even less likely that we could replicate the well-established vulnerability effect in cross-lagged panel models, possibly as a result of more fluctuations in state self-esteem, which would be less predictive. At the same time, such larger fluctuations could allow us to gain further insights when applying statistical models that disentangle state residual and trait variance. Thus, it would be interesting to see whether the first indications that depressive symptoms predict state self-esteem when controlling for the trait variance might be strengthened by employing shorter time frames for self-esteem. Over and above self-esteem and depressive symptoms, we encourage researchers to question the previous traditions that have been applied to measure their constructs of interest, given that the use of certain time frames might be related to the magnitude and direction of the results. However, more research is necessary before clear predictions can be made about whether the effects of certain time frames will hold across a broad variety of constructs.

The Combined Effects of Measuring and Modeling States and Traits in the Reciprocal Analyses between Constructs

Along with making state and trait distinctions by measuring the constructs with different time frames, we separated states and traits statistically in the reciprocal analyses of the two constructs by combining latent state-trait models with cross-lagged panel models. The present study pursued this idea by applying the LST-CLPM to the four experimental conditions that differed in the time frames we used to measure the constructs. On the basis of our results, we see support for three overarching conclusions:

First, across both studies, we observed significant cross-lagged paths in the LST-CLPM only if at least one of the constructs was measured as a state. Across both studies, when both constructs were measured as traits, there were no significant cross-lagged paths. Hence, when using the LST-CLPM with trait measures only, cross-lagged results were rare and therefore not very informative for extending our understanding of the relations between constructs.

Second, at the same time, when one construct was measured as a trait and the other as a state (Conditions A and C), there were somewhat unexpected findings. More specifically, in the confirmatory study, we observed positive cross-lagged paths between self-esteem and depressive symptoms in the two conditions where one construct was measured as a trait and the

other as a state. In both conditions, these positive cross-lagged paths represented the regression coefficient from a trait to a state measure, whereas the respective state measure revealed significant and high autoregressive coefficients. There were indications for this result pattern in the exploratory study as well; however, these paths were not significant. Regarding the relation between self-esteem and depressive symptoms, this result pattern is difficult to interpret, and we cannot rule out statistical explanations for it. It might be possible that an imbalance in the time frames of the two measures fueled these unexpected results given certain modeling strategies. More research using different time frames and different modeling approaches is needed to systematically evaluate such patterns of results.

Third, the use of different time frame measures when using state-trait statistical models points to a specificity that needs more systematic attention: The LST-CLPM estimates cross-lagged paths on the basis of the latent state variables that were used to measure the constructs, while controlling for the trait variance of the respective outcome variable. The more trait variance there was, the less variance remained unexplained in a latent state variable (i.e., latent state residual or time-point-specific deviations), and thus, the less variance there was that could be explained by the autoregressive and cross-lagged paths. On the one hand, the amount of trait and state residual variance depends on the nature of the construct, but on the other hand, we showed that it is also a question of the time frame. Therefore, when researchers are interested in explaining state residual variances (i.e., time-point-specific deviations), we argue that state measures produce the desired and adequate structures (Nezlek, 2007; Podsakoff, Spoelma, Chawla, & Gabriel, 2019). The importance of applying state measures that are more likely to obtain substantial amounts of state residual variance instead of global and highly stable trait measures has been discussed in current personality research in attempts to understand patterns of development as well as their underlying processes (Baumert et al., 2017; Finnigan & Vazire, 2018; Fleeson & Noftle, 2012; Roberts, 2018; Vazire & Sherman, 2017). Our results empirically supported this idea by experimentally showing that there is more state residual variance in state measures, and consequently, more variance can be explained in the LST-CLPM when state measures are used for both constructs.

Advances for Research on the Relation between Self-Esteem and Depressive Symptoms

The present study was effective in analyzing the impact of a two-fold strategy to operationalize states and traits. By applying this idea to the relation between self-esteem and depressive symptoms, we have additionally gained important new insights into this intensively

studied research question. Across two samples, our results further supported the vulnerability model if and only if self-esteem was measured as a trait and the regular CLPM was used. Thereby, our results extend previous empirical evidence by showing that the vulnerability effect is robust (a) over a short time period (3 months), (b) in the present samples (students in their first and second semester), and (c) when depressive symptoms are measured with different time frames. Accordingly, the vulnerability effect seemed to be a very stable finding that indicated that trait self-esteem is associated with future depressive symptoms. However, our results also indicated that when self-esteem was measured as a state, it might be predicted by depressive symptoms. A possible explanation might be that depressive symptoms influence short-term social relationships as well as intrapersonal processes (e.g., selective attention) that can damage a person's state self-esteem. These results suggest mechanisms that follow the assumptions of the scar model in which low self-esteem is a consequence of depressive symptoms.

When controlling for stable trait differences using the LST-CLPM, our results did not support the vulnerability model but instead showed a more diffuse and less consistent pattern of results across the two studies. In the exploratory study, depressive symptoms negatively predicted self-esteem when both constructs were assessed as states, which indicated that individuals' temporary depressive symptoms predicted differences in self-esteem. Even if there were indications for this effect in the confirmatory study, it could not be clearly replicated and should therefore be interpreted cautiously. In the confirmatory study, by contrast, we found several significant cross-paths between self-esteem and depressive symptoms, which to some extent contradicted both the vulnerability and scar models. Across both studies, in the condition that used the traditional measures of self-esteem and depressive symptoms (Condition A: self-esteem trait, depressive symptoms state), our results matched Masselink et al.'s (2018) Study 2 in terms of finding no vulnerability effects but contradicted Masselink et al.'s Studies 1 and 3, which found weak vulnerability effects. However, we used a slightly different analytical approach to control for stable trait differences than Masselink et al. (2018) did. Hence, the results are only partially comparable (see Supplemental Material B6 for the same analytical approach).

In sum, the present investigation supported the vulnerability effect when we employed traditional cross-lagged panel models and used trait self-esteem measures. However, this effect largely did not appear when we used state self-esteem measures. When we controlled for stable trait differences, the findings did not support the vulnerability effect. By contrast, there were

indications that depressive symptoms predicted self-esteem when both constructs were assessed as states; however, drawing conclusions is difficult because the trait-controlled findings did not clearly replicate across the two studies and were somewhat contradictory in the confirmatory study.

Limitations and Future Research

Despite the strengths of our investigation (e.g., we used two studies with a longitudinal, experimental design), which allowed us to apply credible tests of our research questions, some limitations should be kept in mind when interpreting the results. First, the relatively short duration of 3 months should be considered. This was comparably short in contrast to other longitudinal studies, such as some that have also explored the relation between self-esteem and depressive symptoms (Rieger et al., 2016). However, this concern was mitigated by the fact that we replicated prior findings that showed that a trait operationalization of self-esteem predicted state measures of depressive symptoms when applying the traditional CLPM. In addition, the focus of this study was on measurement specificity, and this could be adequately addressed in such a time period. Of course, more measurement occasions across a longer period of time would have been helpful and of substantial interest for assessing both traits and states. This is especially true when the goals are to reliably disentangle state residual and trait variance.

Second, like other previous studies on the relation between self-esteem and depressive symptoms (Orth et al., 2008; Rieger et al., 2016; Sowislo & Orth, 2013), we did not have a clinical sample that included patients with diagnoses such as major depression. Yet, the selection of the sample was driven by the idea that the peak incidence of depression lies in the phase of early adulthood (Kieling et al., 2019). However, our study was not an ideal test of the scar hypothesis. Because we did not assess clinical diagnoses of depression, we could not make any predictions about the consequences of a clinically relevant depressive episode. Nonetheless, the present study provides important considerations about the role of assessment in the analysis of important affective constructs. Clinicians often default to measuring depressive symptoms using time frames that push the construct more strongly in the direction of states than traits, and this will have important implications for its relations with other constructs, such as self-esteem. In this regard, of course, a replication of our design using a clinical sample could provide even a better understanding and could generate even more practical implications.

Third, the present study was effective in showing that the time frame makes a difference in the interplay of self-esteem and depressive symptoms. We provided thoughts about different

processes that underlie the different time frames. Nevertheless, future studies should systematically study these processes (e.g., the different cognitive mechanisms that arise from different time frames). Self-esteem and depressive symptoms have functioned as valid and interesting examples for analyzing the impact of different time frames. But of course this effect needs to be demonstrated with other psychological constructs. Especially when comparing temporal stabilities or analyzing the interplay between psychological constructs, researchers should give thought to the comparability and consequences of the specific time frame and the statistical models they use.

Finally, we want to address the replicability of the results. The present study presented a sophisticated scientific research approach, beginning with an exploratory study followed by a preregistered confirmatory study that challenged the conclusions drawn from the exploratory study. Overall, a large part of the results was replicated in the confirmatory study, especially when pointing to the 95 % confidence intervals of the two studies. Nevertheless, some results were not in line with the hypotheses based on the exploratory study. To further strengthen the evidence on the state-trait differentiation, it would help if other laboratories, preferably multiple laboratories, would embrace our experimental approach and provide a comprehensive set of replication studies in order to increase the independence and generalizability of the results (Hedges & Schauer, 2019; Maxwell, Lau, & Howard, 2015; Simons, 2014). Along with direct replications, such as the one we conducted here, conceptual replications represent a subsequent important evaluation of the impact of measuring and modeling states and traits (Stroebe & Strack, 2014), such as transferring the present design to different constructs, different time frames, or different populations.

Conclusion

Researchers have to decide whether to measure a construct as a state or a trait and whether to apply statistical models that can disentangle state residual and trait variance. In the present experimental investigation, we observed that the measurement and modeling of states and traits is fundamentally related to the results we can expect. The study contributes to the field because we used a two-part strategy to capture states and traits and brought them together to address an intensively studied research question, namely, the nature of the interplay of self-esteem and depressive symptoms. The present research emphasizes the consequences of two different operationalizations of states and traits on the stability of self-esteem and depressive

symptoms as well as on their reciprocal relations. Specifically, the well-researched vulnerability effect, indicating that self-esteem predicted depressive symptoms, only held when trait self-esteem was measured, and when we used regular cross-lagged panel models that did not disentangle state and trait variance. Most important, the results illustrate the need to integrate perspectives on measuring and modeling states and traits in psychological research.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.).* Washington, DC: Author.

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, *28*(11), 1547–1562. https://doi.org/10.1177/0956797617723724

Anderson, S. F., & Maxwell, S. E. (2017). Addressing the "Replication Crisis": Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research, 52*(3), 305–324. https://doi.org/10.1080/00273171.2017.1289361

Asparouhov, T., & Muthen, B. (2008). *Auxiliary variables predicting missing data*. Retrieved from https://www.statmodel.com/download/AuxM2.pdf

Back, M. D., Küfner, A. C. P., Dufner, M., Gerlach, T. M., Rauthmann, J. F., & Denissen, J. J. A. (2013). Narcissistic admiration and rivalry: Disentangling the bright and dark sides of narcissism. *Journal of Personality and Social Psychology, 105*, 1013-1037.

Baumert, A., Schmitt, M., Perugini, M., Johnson, W., Blum, G., Borkenau, P.,. . . Mõttus, R. (2017). Integrating personality structure, personality process, and personality development. *European Journal of Personality*, *31*(5), 503–528. https://doi.org/10.1002/per.2115

Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. New York, NY: Harper & Row.

Bergmann, C., & Eder, F. (2005). *Allgemeiner Interessen-Struktur-Test mit Umwelt-Struktur-Test (UST-R) - Revision.* [General Interest Structure Test and Environmental Structure Test – Revision]. Göttingen, Germany: Beltz.

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

Bleidorn, W. (2012). Hitting the road to adulthood: Short-term personality development during a major life transition. *Personality and Social Psychology Bulletin*, *38*(12), 1594–1608. https://doi.org/10.1177/0146167212456707

Borkenau, P., & Ostendorf, F. (1991). Ein Fragebogen zur Erfassung fünf robuster Persönlichkeitsfaktoren. / A questionnaire for assessing five robust personality factors. *Diagnostica, 37*, 29-41.

Brown, J. D., & Marshall, M. A. (2006). The three faces of self-esteem. In M. H. Kernis (Ed.), *Self-esteem issues and answers. A sourcebook of current perspectives* (pp. 4–9). New York, NY: Psychology Press.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). New York, London: Routledge.

Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, *12*(4), 381–398. https://doi.org/10.1037/1082-989X.12.4.381

Cole, D. A., & Martin, N. C. (2005). The longitudinal structure of the Children's Depression Inventory: Testing a latent trait-state model. *Psychological Assessment*, *17*(2), 144–155. https://doi.org/10.1037/1040-3590.17.2.144

Cole, D. A., Martin, N. C., & Steiger, J. H. (2005). Empirical and conceptual problems with longitudinal trait–state models: Introducing a trait–state–occasion model. *Psychological Methods*, 10, 3–20. doi:10.1037/1082-989X.10.1.3

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*(4), 330–351.

Davitz, J. R. (1969). *The language of emotion*. New York: Academic.

Deinzer, R., Steyer, R., Eid, M., Notz, P., Schwenkmezger, P., Ostendorf, F., & Neubauer, A. (1995). Situational effects in trait assessment: The FPI, NEOFFI, and EPI questionnaires. *European Journal of Personality*, *9*(1), 1–23. https://doi.org/10.1002/per.2410090102

Donnellan, M. B., Kenny, D. A., Trzesniewski, K. H., Lucas, R. E., & Conger, R. D. (2012). Using trait-state models to evaluate the longitudinal consistency of global self-esteem from adolescence to adulthood. *Journal of Research in Personality*, *46*(6), 634–645. https://doi.org/10.1016/j.jrp.2012.07.005

Dumenci, L., & Windle, M. (1996). A latent trait-state model of adolescent depression using the Center for Epidemiologic Studies-Depression Scale. *Multivariate Behavioral Research*, *31*(3), 313–330. https://doi.org/10.1207/s15327906mbr3103_3

Eysenck, H. J. (1983). Cicero and the state-trait theory of anxiety: Another case of delayed recognition. *American Psychologist*, *38*(1), 114–115. https://doi.org/10.1037/0003-066X.38.1.114

Finnigan, K. M., & Vazire, S. (2018). The incremental validity of average state self-reports over global self-reports of personality. *Journal of Personality and Social Psychology*, *115*(2), 321–337. https://doi.org/10.1037/pspp0000136

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*(6), 1011–1027. https://doi.org/10.1037/0022-3514.80.6.1011

Fleeson, W., & Jayawickreme, E. (2015). Whole trait theory. *Journal of Research in Personality*, *56*, 82–92. https://doi.org/10.1016/j.jrp.2014.10.009

Fleeson, W., & Noftle, E. E. (2012). Personality research. In M. R. Mehl, T. S. Conner, & M. Csikszentmihalyi (Eds.), *Handbook of research methods for studying daily life* (pp. 525–538). New York: Guilford Press.

Geiser, C., Götz, T., Preckel, F., & Freund, P. A. (2017). States and traits. *European Journal of Psychological Assessment*, *33*(4), 219–223. https://doi.org/10.1027/1015-5759/a000413

Gerlitz, J-Y., & Schupp, J. (2005). *Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im Sozio-ökonomischen Panel (SOEP).* [Assessment of Big-Five-based personality traits in the Socioeconomic Panel (SOEP).]. Berlin: Deutsches Institut für Wirtschaftsforschung.

Geukes, K., Nestler, S., Hutteman, R., Dufner, M., Küfner, A. C. P., Egloff, B.,. . . Back, M. D. (2017). Puffed-up but shaky selves: State self-esteem level and variability in narcissists. *Journal of Personality and Social Psychology*, *112*(5), 769–786. https://doi.org/10.1037/pspp0000093

Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*(1), 80–100. https://doi.org/10.1207/S15328007SEM1001_4

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

Hautzinger, M., & Bailer, M. (1993). *Allgemeine Depressions Skala - ADS [Center for Epidemiologic Studies–Depression Scale, CES-D]*. Weinheim: Beltz.

Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology*, *60*(6), 895–910. https://doi.org/10.1037/0022-3514.60.6.895

Hedges, L. V., & Schauer, J. M. (2019). More than one replication study is needed for unambiguous tests of replication. *Journal of Educational and Behavioral Statistics*, *44*(5), 543–570. https://doi.org/10.3102/1076998619852953

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Howell, R. T., Ksendzova, M., Nestingen, E., Yerahian, C., & Iyer, R. (2017). Your personality on a good day: How trait and state personality predict daily well-being. *Journal of Research in Personality*, *69*, 250–263. https://doi.org/10.1016/j.jrp.2016.08.001

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

Jobe, J. B., Tourangeau, R., & Smith, A. F. (1993). Contributions of survey research to the understanding of memory. *Applied Cognitive Psychology*, *7*(7), 567–584. https://doi.org/10.1002/acp.2350070703

Kandler, C., Zimmermann, J., & McAdams, D. P. (2014). Core and surface characteristics for the description and theory of personality differences and development. *European Journal of Personality*, *28*(3), 231–243. https://doi.org/10.1002/per.1952

Kenny, D. A., & Zautra, A. (1995). The trait-state-error model for multiwave data. *Journal of Consulting and Clinical Psychology*, *63*(1), 52–59. https://doi.org/10.1037/0022-006X.63.1.52

Kenny, D. A., & Zautra, A. (2001). Trait–state models for longitudinal data. In L. M. Collins (Ed.), *Decade of behavior. New methods for the analysis of change* (pp. 243–263). Washington D.C.: American Psychological Association. https://doi.org/10.1037/10409-008

Kernis, M. H., Cornell, D. P., Sun, C.-R., Berry, A., & Harlow, T. (1993). There's more to self-esteem than whether it is high or low: The importance of stability of self-esteem. *Journal of Personality and Social Psychology*, *65*(6), 1190–1204. https://doi.org/10.1037/0022-3514.65.6.1190

Kernis, M. H. (Ed.). (2006). *Self-esteem issues and answers: A sourcebook of current perspectives*. New York, NY: Psychology Press.

Kieling, C., Adewuya, A., Fisher, H. L., Karmacharya, R., Kohrt, B. A., Swartz, J. R., & Mondelli, V. (2019). Identifying depression early in adolescence. *The Lancet Child & Adolescent Health*, *3*(4), 211–213. https://doi.org/10.1016/S2352-4642(19)30059-8

Kovaleva, A., Beierlein, C., Kemper, C. J., & Rammstedt, B. (2012a). *Eine Kurzskala zur Messung von Impulsivität nach dem UPPS-Ansatz: Die Skala Impulsives Verhalten (I-S)* [A short scale for assessing impulsivity following the UPPS approach]. GESIS Working Papers 2012|20. Köln: GESIS.

Kovaleva, A., Beierlein, C., Kemper, C. J., & Rammstedt, B. (2012b). *Eine Kurzskala zur Messung von Kontrollüberzeugung: Die Skala Internale-Externale-Kontrollüberzeugung (IE-4)* [A short scale for assessing control beliefs: The scale Internal-External Control Beliefs-4 (IE-4)]. GESIS Working Papers 2012|19. Köln: GESIS.

Lance, C. E., Christie, J., & Williamson, G. M. (2019). Do State and Trait Measures Measure States and Traits? The Case of Community-Dwelling Caregivers of Older Adults. *Assessment*, 1073191119888582. https://doi.org/10.1177/1073191119888582

Lewinsohn, P. M., Steinmetz, J. L., Larson, D. W., & Franklin, J. (1981). Depression-related cognitions: Antecedent or consequence? *Journal of Abnormal Psychology*, *90*(3), 213–219. https://doi.org/10.1037/0021-843X.90.3.213

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 151–173. https://doi.org/10.1207/S15328007SEM0902_1

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

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(1), 59–72. https://doi.org/10.1207/s15328007sem1301_3

MacCallum, R. C., Browne, M. W., & Cai, L. (2006). Testing differences between nested covariance structure models: Power analysis and null hypotheses. *Psychological Methods*, *11*(1), 19–35. https://doi.org/10.1037/1082-989X.11.1.19

Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J. S., & Davier, M. von. (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*(3), 257–284. https://doi.org/10.1037/a0032773

Masselink, M., van Roekel, E., Hankin, B. L., Keijsers, L., Lodder, G.M.A., Vanhalst, J.,. . . Laceulle, O. (2018). The longitudinal association between self-esteem and depressive symptoms in adolescents: Separating between-person effects from within-person effects. *European Journal of Personality*, *96*, 35. https://doi.org/10.1002/per.2179

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist*, *70*(6), 487–498. https://doi.org/10.1037/a0039400

McCrae, R. R., & Costa, P. T. (2008). The five-factor theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality. Theory and research* (pp. 159–181). New York, NY: Guilford Press.

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology*, *4*(1), 103–120. https://doi.org/10.1080/0022250X.1975.9989847

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525-543.

Mischel, W. (2004). Toward an integrative science of the person. *Annual Review of Psychology*, *55*, 1–22. https://doi.org/10.1146/annurev.psych.55.042902.130709

Mohiyeddini, C., Hautzinger, M., & Bauer, S. (2002). Eine Latent-State-Trait-Analyse zur Bestimmung der dispositionellen und zustandsbedingten Anteile dreier Instrumente zur Erfassung von Depressionen: ADS, BDI und SDS [A latent state-trait analysis on assessing trait and state components of three instruments for measuring depression: ADS, BDI, and SDS]. *Diagnostica*, *48*(1), 12–18. https://doi.org/10.1026//0012-1924.48.1.12

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide (8th ed.)*. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2002). How to use a monte carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8

Neubauer, A. B., Scott, S. B., Sliwinski, M. J., & Smyth, J. M. (2019). How was your day? Convergence of aggregated momentary and retrospective end-of-day affect ratings across the adult life span. *Journal of Personality and Social Psychology.* Advance online publication. https://doi.org/10.1037/pspp0000248

Nezlek, J. B. (2007). A multilevel framework for understanding relationships among traits, states, situations and behaviours. *European Journal of Personality*, *21*(6), 789–810. https://doi.org/10.1002/per.640

Ormel, J., Oldehinkel, A. J., & Vollebergh, W. (2004). Vulnerability before, during, and after a major depressive episode: a 3-wave population-based study. *Archives of General Psychiatry*, *61*(10), 990–996. https://doi.org/10.1001/archpsyc.61.10.990

Orth, U., Robins, R. W., Trzesniewski, K. H., Maes, J., & Schmitt, M. (2009). Low self-esteem is a risk factor for depressive symptoms from young adulthood to old age. *Journal of Abnormal Psychology*, *118*(3), 472–478. https://doi.org/10.1037/a0015922

Orth, U., Robins, R. W., & Meier, L. L. (2009). Disentangling the effects of low self-esteem and stressful events on depression: Findings from three longitudinal studies. *Journal of Personality and Social Psychology*, *97*(2), 307–321. https://doi.org/10.1037/a0015645

Orth, U., Robins, R. W., & Roberts, B. W. (2008). Low self-esteem prospectively predicts depression in adolescence and young adulthood. *Journal of Personality and Social Psychology*, *95*(3), 695–708. https://doi.org/10.1037/0022-3514.95.3.695

Podsakoff, N. P., Spoelma, T. M., Chawla, N., & Gabriel, A. S. (2019). What predicts within-person variance in applied psychology constructs? An empirical examination. *The Journal of Applied Psychology*, *104*(6), 727–754. https://doi.org/10.1037/apl0000374

Pöhlmann, K., & Brunstein, J. C. (1997). Ein Fragebogen zur Messung von Lebenszielen [A questionnaire for the assessment of life goals]. *Diagnostica, 43,* 63–79.

Prenoveau, J. M. (2016). Specifying and Interpreting Latent State–Trait Models With Auto-regression: An Illustration. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(5), 731-749. https://doi.org/10.1080/10705511.2016.1186550

Radloff, L. S. (1977). The CES-D Scale: A self report depression scale for research in the general population. *Applied Psychological Measurement, 1*(3), 385–401. https://doi.org/10.1177/014662167700100306

Ravens-Sieberer, U., Gosch, A., Abel, T., Auquier, P., Bellach, B.-M., Bruil, J.,. . . Rajmil, L. (2001). Quality of life in children and adolescents: A European public health perspective. *Sozial- und Präventivmedizin SPM*, *46*(5), 294–302. https://doi.org/10.1007/BF01321080

Rieger, S., Göllner, R., Spengler, M., Trautwein, U., Nagengast, B., & Roberts, B. W. (2017). Social cognitive constructs are just as stable as the big five between grades 5 and 8. *AERA Open*, *3*(3), 233285841771769. https://doi.org/10.1177/2332858417717691

Rieger, S., Göllner, R., Trautwein, U., & Roberts, B. W. (2016). Low self-esteem prospectively predicts depression in the transition to young adulthood: A replication of Orth, Robins, and Roberts (2008). *Journal of Personality and Social Psychology*, *110*(1), e16-e22. https://doi.org/10.1037/pspp0000037

Roberts, B. W. (2018). A revised sociogenomic model of personality traits. *Journal of Personality*, *86*, 23–35. https://doi.org/10.1111/jopy.12323

Roberts, B. W., & Jackson, J. J. (2008). Sociogenomic personality psychology. *Journal of Personality*, *76*(6), 1523–1544. https://doi.org/10.1111/j.1467-6494.2008.00530.x

Robinson, M. D., & Clore, G. L. (2002). Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes. *Journal of Personality and Social Psychology*, *83*(1), 198–215. https://doi.org/10.1037/0022-3514.83.1.198

Rosenberg, M. (1979). *Conceiving the self*. New York: Basic Books.

Rosenberg, M. (1989). *Society and the adolescent self-image*. Middletown, CT: Wesleyan University Press.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*(4), 507–514. https://doi.org/10.1007/BF02296192

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *9*(1), 76–80. https://doi.org/10.1177/1745691613514755

Sowislo, J. F., & Orth, U. (2013). Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies. *Psychological Bulletin*, *139*(1), 213–240. https://doi.org/10.1037/a0028931

Sowislo, J. F., Orth, U., & Meier, L. L. (2014). What constitutes vulnerable self-esteem? Comparing the prospective effects of low, unstable, and contingent self-esteem on depressive symptoms. *Journal of Abnormal Psychology*, *123*(4), 737–753. https://doi.org/10.1037/a0037770

Spaderna, H., Schmukle, S. C., & Krohne, H. W. (2002). Bericht über die deutsche Adaptation der State-Trait Depression Scales (STDS) [Report about the German adaptation of State-Trait Depression Scales (STDS)]. *Diagnostica*, *48*(2), 80–89. https://doi.org/10.1026//0012-1924.48.2.80

Spielberger, C. D. (1966). *Anxiety and Behavior*. New York: Academic Press.

Spielberger, C. D. (1995). *State-Trait Depression Scales (Form X-1)*. Palo Alto, CA: Mind Garden.

Spielberger, C. D., Ritterband, L. M., Reheiser, E. C., & Brunner, T. M. (2003). The nature and measurement of depression. *International Journal of Clinical and Health Psychology*, *3*(2), 209–234.

Steyer, R., & Schmitt, T. (1994). The theory of confounding and its application in causal modeling with latent variables. In A. von Eye & C. C. Clogg (Eds.), *Latent variable analysis: Applications for developmental research* (pp. 36–67). Thousand Oaks, CA: Sage.

Steyer, R., Mayer, A., Geiser, C., & Cole, D. A. (2015). A theory of states and traits-revised. *Annual Review of Clinical Psychology*, *11*, 71–98. https://doi.org/10.1146/annurev-clinpsy-032813-153719

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality*, *13*(5), 389–408. https://doi.org/10.1002/(SICI)1099-0984(199909/10)13:5<389::AID-PER361>3.0.CO;2-A

Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *9*(1), 59–71. https://doi.org/10.1177/1745691613514450

Stumpf, H., Angleitner, A., Wieck, T., Jackson, D.N., & Beloch-Till, H. (1985). *Deutsche Personality Research Form (PRF) Handanweisung.* [German Personality Research Form (PRF)]. Göttingen: Hogrefe.

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2009). *The psychology of survey response*. Cambridge: Cambridge Univ. Press.

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

Vazire, S., & Sherman, R. A. (2017). Introduction to the special issue on within-person variability in personality. *Journal of Research in Personality*, *69*, 1–3. https://doi.org/10.1016/j.jrp.2017.07.004

Wagner, J., Lüdtke, O., & Trautwein, U. (2016). Self-esteem is mostly stable across young adulthood: Evidence from Latent STARTS Models. *Journal of Personality*, *84*(4), 523–535. https://doi.org/10.1111/jopy.12178

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Webster, G. D., Smith, C. V., Brunell, A. B., Paddock, E. L., & Nezlek, J. B. (2017). Can Rosenberg's (1965) Stability of Self Scale capture within-person self-esteem variability? Meta-analytic validity and test–retest reliability. *Journal of Research in Personality*, *69*, 156–169. https://doi.org/10.1016/j.jrp.2016.06.005

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, *4*(1), 10-18.

Zillig, L. M. P., Hemenover, S. H., & Dienstbier, R. A. (2002). What do we assess when we assess a big 5 trait? A content analysis of the affective, behavioral, and cognitive processes represented in big 5 personality inventories. *Personality and Social Psychology Bulletin*, *28*(6), 847–858. https://doi.org/10.1177/0146167202289013

Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z.,. . . Diener, E. (2019). From data to causes I: Building a general cross-lagged panel model (GCLM). *Organizational Research Methods*, *175*(4), 109442811984727. https://doi.org/10.1177/1094428119847278

Zyphur, M. J., Voelkle, M. C., Tay, L., Allison, P. D., Preacher, K. J., Zhang, Z.,. . . Diener, E. (2019). From data to causes II: Comparing approaches to panel data analysis. *Organizational Research Methods*, *101*(2), 109442811984728. https://doi.org/10.1177/1094428119847280

Supplemental Material

Supplemental Material A: Method

Supplemental Material A1: Attrition Analyses

Table A1
Attrition Analysis: Differences in Relevant Outcomes Variables

	Exploratory study			Confirmatory study		
Dependent variable	*N*	*d*	*p*	*N*	*d*	*p*
High School GPA	636	0.363	.000	1004	0.153	.007
Education mother	591	-0.163	.135	949	0.082	.348
Education father	581	0.048	.664	922	0.046	.610
SE-Trait T1	326	-0.178	.198	519	0.038	.744
SE-State T1	322	-0.314	.030	517	-0.120	.392
Dep-Trait T1	313	0.302	.035	515	0.109	.369
Dep-State T1	330	0.026	.853	508	-0.164	.121
Openness T1	648	0.008	.939	1037	-0.064	.438
Conscientiousness T1	648	-0.603	.000	1037	-0.327	.000
Extraversion T1	648	-0.012	.909	1037	0.067	.417
Agreeableness T1	648	-0.046	.645	1037	-0.033	.689
Neuroticism T1	648	0.094	.349	1037	0.042	.610
Gender (female = 1)	648	*OR* = 0.797	.322	1031	*OR* = 0.666	.020

Note. d = standardized mean difference; independent variable = 1 indicates missing at a minimum of one time point; higher GPA indicates lower achievement.

Supplemental Material A2: Additional Measures Included in the Data Sets

Table A2
Additional Measures Included in the Data Set Used for the Exploratory Study

Measure	Number of items	Time points
NEO Five Factor Inventory (Borkenau & Ostendorf, 1991)	60	T1, T4
Big Five Inventory – Short (Gerlitz & Schupp, 2005)	15	T1, T2, T3, T4
Revised General Interest Structure Test (Bergmann & Eder, 2005)	60	Distributed across the four time points
New items on vocational interests (own construction)	42	T4
Achievement motivation and goals (Stumpf, Angleitner, Wieck, Jackson, & Beloch-Till, 1985)	16	T1, T2, T3, T4
Time spent on studying (Bleidorn, 2012)	3	T1, T2, T3, T4
Future plans (Pöhlmann & Brunstein, 1997)	8	T1, T4
Impulsive behavior (Kovaleva, Beierlein, Kemper, & Rammstedt, 2012a)	8	T1
Control beliefs (Kovaleva, Beierlein, Kemper, & Rammstedt, 2012b)	4	T1
Narcissistic Admiration and Rivalry Questionnaire– State/Trait (Back, Küfner, Dufner, Gerlach, Rauthmann, & Denissen, 2013)	6	T1, T2, T3, T4
Subjective health	2	T4
Demographic variables	11	T1

Table A3
Additional Measures Included in the Data Set Used for the Confirmatory Study

Measure	Number of items	Time points
NEO Five Factor Inventory (Borkenau & Ostendorf, 1991)	60	T1, T4 (at T4 only 6 items)
Effort (own development)	8	T1, T2, T3, T4
Interest in the academic major (own development)	3	T1, T4
Demographic variables	11	T1

Supplemental Material A3: Instruments

To measure self-esteem and depressive symptoms as both a state and a trait, we varied the time frame in the questionnaires ("In general" vs. "During the last 2 weeks). The traditional self-esteem measure was the 10-item Rosenberg Self-Esteem Scale (Rosenberg, 1989), which assesses self-esteem as a trait. The items are displayed in Table A4. The traditional depressive symptoms measure was the 15-item Center for Epidemiological Studies Depression Scale (Radloff, 1977), which reflects a state measure. For both measures, we present standardized factor loadings of single-factor confirmatory factor analyses for all four time points for both studies (see Tables A4 and A5). For all analyses, we constructed three random item parcels for both self-esteem (Parcel 1: Item 2, Item 7, Item 9, Item 10; Parcel 2: Item 1, Item 4, Item 8; Parcel 3: Item 3, Item 5, Item 6) and depressive symptoms (Parcel 1: Item 1, Item 3, Item 9, Item 11, Item 14; Parcel 2: Item 2, Item 4, Item 7, Item 12, Item 15; Parcel 3: Item 5, Item 6, Item 8, Item 10, Item 13).

In order to assess self-esteem as a state and depressive symptoms as a trait, the tense of the verbs in all items had to be adapted (e.g., self-esteem as a trait: "I am satisfied with myself"; self-esteem as a state: "I was satisfied with myself"). Moreover, for three items, the original item wording had to be adapted so that the items would make logical sense with the new time frame: (a) self-esteem trait: "I think I am no good at all"; self-esteem state: "I often thought I was no good"; (b) depressive symptoms trait: "I talk less than in the past"; depressive symptoms state: "I talked less than usual"; (c) depressive symptoms trait: "I am bothered by things more than in the past"; depressive symptoms state: "I was bothered by things that usually don't bother me"). In order to control for the impact of these three items, we repeated the analyses without the items in the respective conditions for both studies. The results were virtually unchanged. However, in the exploratory study in the CLPM in Condition D, the paths from depressive symptoms to self-esteem were only significant for a one-tailed test (.17 [-.38, .03], -.18 [-.40, .03], -.18 [-.39, .03]). In the confirmatory study, in the CLPM in Condition D, the paths from self-esteem to depressive symptoms were significant for a one-tailed test (.13 [-.28, .03], -.13 [-.28, .03], -.12 [-.27, .02]), and in the LST-CLPM in Condition C, for both directions, the cross-lagged paths were only significant for a one-tailed test (-.18 [-.34, .00], -.17 [-.34, .00], -.16 [-.33, .01], .21 [-.02, .43], .22 [-.01, .46], .21 [-.01, .43]).

Table A4
Self-Esteem Items (Rosenberg, 1989)

No.	Self-esteem trait	Factor loadings *E: (T1, T2, T3, T4)* *C: (T1, T2, T3, T4)*	Self-esteem state	Factor loadings *E: (T1, T2, T3, T4)* *C: (T1, T2, T3, T4)*
	In general ...		**During the last 2 weeks ...**	
1	...I am satisfied with myself.	*E: .76, .79, .72, .77* *C: 74, .72, .69, .76*	...I was satisfied with myself.	*E: .72, .73, .73, .74* *C: .73, .66, .67, .70*
2	...I think I am no good at all. (R)	*E: .76, .78, .76, .81* *C: .81, .78, .82, .83*	...I often thought I was no good. (R)	*E: .83, .80, .77, .85* *C: .83, .80, .79, .80*
3	...I feel that I have a number of good qualities.	*E: .59, .65, .66, .66* *C: .47, .58, .63, .60*	...I felt I had a number of good qualities.	*E: .54, .57, .68, .69* *C: .57, .61, .61, .60*
4	...I am able to do things as well as most other people.	*E: .49, .53, .55, .65* *C: .44, .51, .61, .59*	...I was able to do things as well as most other people.	*E: .53, .57, .65, .68* *C: .51, .64, .64, .63*
5	...I feel I do not have much to be proud of. (R)	*E: .73, .75, .72, .77* *C:.73, .79, .77, .74*	...I felt that I did not have much to be proud of. (R)	*E: .74, .74, .80, .82* *C: .76, .78, .77, .81*
6	...I certainly feel useless at times. (R)	*E:.77, .75, .75, .80* *C: .80, .76, .80, .81*	…I certainly felt useless at times. (R)	*E: .81, .79, .79, .82* *C: .82, .79, .78, .78*
7	...I feel that I'm a person of worth, at least on an equal plane with others.	*E: .72, .77, .74, .71* *C: .66, .68, .71, .70*	...I felt that I was a person of worth, at least on an equal plane with others.	*E: .58, .60, .71, .69* *C: .67, .67, .69, .67*
8	…I wish I could have more respect for myself. (R)	*E: .71, .66, .70, .68* *C: .68, .66, .67, .65*	…I wished I could have more respect for myself. (R)	*E: .55, .65, .60, .55* *C: .63, .57, .64, .67*
9	… I am inclined to feel that I am a failure. (R)	*E: .78, .80, .82, .84* *C: .82, .82, .82, .83*	… I was inclined to feel that I am a failure. (R)	*E: .83, .78, .85, .88* *C: .84, .84, .83, .84*
10	… I take a positive attitude toward myself.	*E: .83, .84, .81, .80* *C: .82, .79, .81, .82*	… I took a positive attitude toward myself.	*E: .71, .75, .75, .79* *C: .79, .72, .75, .71*

Note. E = Exploratory Study, C = Confirmatory Study; Items reproduced with permission from Wesleyan University Press. For further information on the German version of the items, please contact the authors.

Table A5

Depressive Symptoms Items (Hautzinger & Bailer, 1993; Radloff, 1977)

No.	No. in Radloff (1977)	Depressive symptoms trait: Factor loadings *E: (T1, T2, T3, T4)* *C: (T1, T2, T3, T4)*	Depressive symptoms state: Factor loadings *E: (T1, T2, T3, T4)* *C: (T1, T2, T3, T4)*
1	1	*E:.43, .48, .55, .49* *C: .40, .50, .49, .55*	*E: .42, .61, .61, .57* *C: .44, .50, .58, .58*
2	3	*E: .58, .66, .70, .73* *C: .67, .74, .77, .76*	*E:.76, .80, .81, .80* *C: .75, .79, .76, .78*
3	5	*E: .46, .52, .49, .46* *C: .48, .50, .55, .55*	*E: .54, .56, .56, .62* *C: .42, .51, .46, .44*
4	6	*E: .82, .83, .84, .87* *C: .85, .87., 86., 86*	*E: .83, .83, .86, .82* *C: .83, .85, .86, .85*
5	7	*E:.63, .73, .74, .71* *C: .65, .68, .74, .72*	*E: .62, .64, .71, .64* *C: .61, .62, .63, .61*
6	9	*E: .69, .69, .73, .65* *C: .73, .72, .70, .72*	*E: .64, .64, .69, .71* *C: .66, .67, .67, .66*
7	10	*E: .66, .69, .69, .71* *C: .66, .68, .70, .75*	*E: .61, .65, .64, .61* *C: .55, .61, .64, .62*
8	11	*E: .40, .53, .53, .53* *C: .51, .55, .51, .59*	*E: .47, .49, .50, .54* *C: .42, .47, .43, .49*
9	12	*E: .68, .69, .68, .76* *C: .66, .74, .67, .73*	*E: .69, .70, .74, .74* *C:.68, .66, .72, .74*
10	13	*E: .39, .44, .50, .34* *C: .41, .48, .50, .54*	*E: .39, .57, .53, .55* *C: .43, .46, .58, .56*
11	14	*E: .71, .68, .68, .67* *C: .64, .73, .72, .71*	*E: .66, .64, .70, .64* *C: .61, .59, .65, .66*
12	16	*E: .73, .69, .73, .68* *C: .67, .74, .67, .69*	*E: .65, .67, .70, .59* *C: .67, .60, .65, .66*
13	18	*E: .82, .80, .81, .83* *C: .81, .82, .82, .85*	*E: .77, .75, .74, .76* *C:.79, .75, .78, .78*
14	19	*E: .51, .56, .54, .50* *C: .56, .52, .55, .55*	*E: .54, .57, .54, .51* *C: .43, .50, .54, .55*
15	20	*E: .54, .54, .56, .55* *C: .54, .66, .63, .62*	*E: .45, .58, .54, .54* *C: .48, .52, .50, .50*

Note. E = Exploratory Study, C = Confirmatory Study. For further information on the German or English version of the items, please contact the authors.

Supplemental Material A4: Differences in the Data Collection and the Compositions of the Samples used in the Exploratory and Confirmatory Studies

Table A6
Differences in the Data Collection and the Compositions of the Samples used in the Exploratory and Confirmatory Studies

	Exploratory study	Confirmatory study
Recruitment of participants	First semester lectures in the Department of Economics and Social Sciences and the Department of Science of the University of Tübingen	Mostly second semester lectures in all departments of the University of Tübingen
Incentive structure	Feedback on the interest and personality profiles; 20 Euro for every participant	Feedback on the personality profile; lottery for 3 iPads and 30x30 Euro gift cards
Time period for data collection	November 2016 - February 2017	May 2019 - July 2019
Length of the online survey (for details, see supplemental material C)	T1, T4: 25 min T2, T3: 5 min	T1: 10 min T2-T4: 2-3 min
Sample size	Total N = 683 (N ~ 170 per condition at T1)	Total N = 1,087 (N ~ 250 per condition at T1)
Age of the sample at T1	M = 20.34, SD = 2.89	M = 21.52, SD = 3.34
Students' majors in the sample (largest groups)	Psychology (25.5%), Economics (14.8%), Teacher training (7.3%)	Medicine (12.3 %), Law (12.0 %), Education (10.7 %)
Dropout rate (dropped out through the course of the study)	18.8%	21.5%

Note. Dropout rate does not include those who joined the study late.

Supplemental Material B: Additional Results

Supplemental Material B1: Measurement Invariance and Model Fit

Without establishing measurement invariance, we would not be able to rule out the possibility that differences in the correlations and regression coefficients (over time) were due to measurement artifacts (Meredith, 1993; Widaman, Ferrer, & Conger, 2010). For this purpose, we specified three models by beginning with a model in which configural measurement invariance was imposed over time (i.e., we tested for the dimensionality of the construct over time). In the next step, we imposed weak measurement invariance over time (i.e., the same factor loadings for the same items). In the last step, we imposed strong measurement invariance over time (i.e., the same factor loadings and intercepts for the same items). Longitudinal constraints were set within conditions, and the models presented here included all four conditions simultaneously. Tables B1 and B2 present goodness-of-fit indices for the measurement models. In both studies, all three models showed a very good fit in terms of the classical fit indices such as the CFI (≥ .989), the TLI (≥ .986), the RMSEA (≤ .034), and the SRMR (≤ .043). Although additional restrictions in the strong invariance model led to a significantly worse Chi-Square (exploratory study: $\Delta\chi^2 = 75.71$, $\Delta df = 48$; $\Delta p = .01$; confirmatory study: $\Delta\chi^2 = 104.03$, $\Delta df = 48$; $\Delta p = .00$), we assumed strong invariance in further analyses because all other fit indices did not decrease substantially (< .01), and the models fit the data very well. In detail, when analyzing the invariance separately for the constructs and for both the trait and the state versions, we found strong invariance for all measures in both studies (see Tables B3 and B4). In both studies, for the state measures of depressive symptoms, and in the confirmatory study for the depressive symptoms trait measure and the self-esteem state measure, although additional restrictions in the strong invariance model led to a significantly worse Chi-Square (see Table B3 an B4), we assumed strong invariance in these measures because most of the fit indices did not decrease substantially (< .01), and the models fit the data very well.

Tables B1 and B2 also present the goodness of fit of the structural models. We compared a cross-lagged panel model that had freely estimated regression coefficients (Model 4) with a model that had longitudinal constraints on all regression coefficients (Model 5). As indicated in Table B1, in the exploratory study, the model fit was good in both Model 4 ($\chi^2 = 1146.95$, $df = 896$, CFI = .984, TLI = .980, RMSEA = .041, SRMR = .060) and Model 5 ($\chi^2 = 1186.30$, $df = 928$, CFI = .983, TLI = .980, RMSEA = .040, SRMR = .067), and the models did not differ significantly ($\Delta\chi^2 = 39.54$, $\Delta df = 32$, $\Delta p = 0.17$). Likewise in the confirmatory study,

as indicated in Table B2, the model fit was good in both Model 4 (χ^2 = 1276.48, *df* = 896, CFI = .984, TLI = .981, RMSEA = .040, SRMR = .051) and Model 5 (χ^2 = 1334.98, *df* = 928, CFI = .983, TLI = .980, RMSEA = .040, SRMR = .055). The models differed significantly ($\Delta\chi^2$ = 56.98, Δdf = 32, Δp = 0.01), but none of the fit indices decreased substantially (< .01). Consequently, in both studies, we focused on the more restrictive model (Model 5). In the next step, we compared a latent state-trait cross-lagged panel model that had freely estimated regression coefficients (Model 6) with a model that had longitudinal constraints on all regression coefficients (Model 7). As indicated in Table B1, in the exploratory study, the model fit was good in both Model 6 (χ^2 = 1093.56, *df* = 884, CFI = .986, TLI = .983, RMSEA = .037, SRMR = .047) and Model 7 (χ^2 = 1136.24, *df* = 916, CFI = .986, TLI = .983, RMSEA = .038, SRMR = .057), and the models did not differ significantly ($\Delta\chi^2$ = 42.45, Δdf = 32, Δp = .10). Likewise, in the confirmatory study, as indicated in Table B2, the model fit was good in both Model 6 (χ^2 = 1181.16, *df* = 884, CFI = .988, TLI = .985, RMSEA = .035, SRMR = .038) and Model 7 (χ^2 = 1207.78, *df* = 916, CFI = .988, TLI = .985, RMSEA = .034, SRMR = .042), and the models did not differ significantly ($\Delta\chi^2$ = 29.99, Δdf = 32, Δp = .57).Consequently, in both studies, we focused on the more restrictive model (Model 7). However, the results for the cross-lagged panel model and the latent state-trait intercept cross-lagged panel model based on the less restrictive model with freely estimated regression coefficients (Model 4, Model 6) are presented in Tables B9 and B10.

Table B1
Fit Indices for the Models Tested in the Exploratory Study

	χ^2	*df*	SF	CFI	TLI	RMSEA	SRMR	$\Delta\chi^2$	Δdf	Δp
Measurement models										
1. Free loadings	887.69	752	1.032	0.991	0.987	0.023	0.034			
2. Longitudinal constraints on loadings	939.77	800	1.032	0.991	0.987	0.032	0.042	52.08	48	0.32
3. Longitudinal constraints on loadings and intercepts	1014.85	848	1.030	0.989	0.986	0.034	0.043	75.71	48	0.01
Structural models										
4. CLPM - Free structural coefficients	1146.95	896	1.027	0.984	0.980	0.041	0.060			
5. CLPM - Longitudinal constraints on regression coefficients	1186.30	928	1.032	0.983	0.980	0.040	0.067	39.54	32	0.17
6. LST-CLPM - Free structural coefficients	1093.56	884	1.024	0.986	0.983	0.037	0.047			
7. LST-CLPM - Longitudinal constraints on regression coefficients	1136.24	916	1.027	0.986	0.983	0.038	0.057	42.45	32	0.10

Note. N = 683. Results based on the MLR estimator (Muthén, & Muthén, 1998-2017); SF = Scaling Factor. CFI = comparative fit index; TLI = Tucker Lewis Index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual. All conditions were estimated simultaneously within one model.

Table B2
Fit Indices for the Models Tested in the Confirmatory Study

	χ^2	*df*	SF	CFI	TLI	RMSEA	SRMR	$\Delta\chi^2$	Δdf	Δp
Measurement models										
1. Free loadings	938.45	752	1.051	0.992	0.989	0.030	0.027			
2. Longitudinal constraints on loadings	992.98	800	1.046	0.992	0.989	0.030	0.033	54.10	48	0.25
3. Longitudinal constraints on loadings and intercepts	1094.88	848	1.043	0.990	0.987	0.033	0.035	104.03	48	0.00
Structural models										
4. CLPM - Free structural coefficients	1276.48	896	1.045	0.984	0.981	0.040	0.051			
5. CLPM - Longitudinal constraints on regression coefficients	1334.98	928	1.050	0.983	0.980	0.040	0.055	56.98	32	0.00
6. LST-CLPM - Free structural coefficients	1181.16	884	1.038	0.988	0.985	0.035	0.038			
7. LST-CLPM - Longitudinal constraints on regression coefficients	1207.78	916	1.048	0.988	0.985	0.034	0.042	29.99	32	0.57

Note. N = 1,087. Results based on the MLR-estimator (Muthén, & Muthén, 1998-2017); SF = Scaling Factor. CFI = comparative fit index; TLI = Tucker Lewis Index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual. All conditions were estimated simultaneously within one model.

Table B3
Measurement Invariance over Time Presented Separately for Each Instrument in the Exploratory Study

	χ^2	*df*	SF	CFI	TLI	RMSEA	SRMR	$\Delta\chi^2$	Δdf	Δp
Self-esteem state										
1a. Free loadings	45.04	30	1.121	0.994	0.988	0.038	0.025			
2a. Constraints on loadings	53.63	36	1.109	0.994	0.988	0.038	0.044	8.17	6	0.226
3a. Constraints on loadings and intercepts	65.63	42	1.093	0.991	0.986	0.041	0.045	11.16	6	0.085
Self-esteem trait										
1b. Free loadings	32.72	30	1.114	0.999	0.998	0.016	0.013			
2b. Constraints on loadings	36.87	36	1.087	1.000	1.000	0.008	0.025	3.81	6	0.702
3c. Constraints on loadings and intercepts	39.91	42	1.075	1.000	1.001	0.000	0.023	2.82	6	0.831
Depressive symptoms state										
1c. Free loadings	30.67	30	1.121	1.000	1.000	0.008	0.022			
2c. Constraints on loadings	38.97	36	1.107	0.999	0.998	0.015	0.028	8.45	6	0.207
3c. Constraints on loadings and intercepts	62.01	42	1.095	0.993	0.989	0.037	0.034	24.20	6	0.000
Depressive symptoms trait										
1d. Free loadings	50.77	30	1.063	0.994	0.988	0.046	0.024			
2d. Constraints on loadings	57.11	36	1.063	0.994	0.989	0.042	0.029	6.34	6	0.386
3d. Constraints on loadings and intercepts	62.91	42	1.056	0.994	0.991	0.039	0.029	5.71	6	0.456

Note. N = 683. Results based on the MLR estimator (Muthén, & Muthén, 1998-2017). SF = Scaling Factor; CFI = comparative fit index; TLI = Tucker Lewis Index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

Table B4
Measurement Invariance over Time Presented Separately for Each Instrument in the Confirmatory Study

	χ^2	*df*	SF	CFI	TLI	RMSEA	SRMR	$\Delta\chi^2$	Δdf	Δp
Self-esteem state										
1a. Free loadings	27.48	30	1.192	1.000	1.001	0.000	0.015			
2a. Constraints on loadings	36.90	36	1.157	1.000	1.000	0.007	0.032	10.12	6	0.120
3a. Constraints on loadings and intercepts	57.13	42	1.135	0.997	0.995	0.026	0.032	22.08	6	0.001
Self-esteem trait										
1b. Free loadings	39.35	30	1.084	0.998	0.996	0.024	0.013			
2b. Constraints on loadings	46.25	36	1.058	0.998	0.997	0.023	0.022	6.76	6	0.343
3c. Constraints on loadings and intercepts	54.02	42	1.050	0.998	0.997	0.023	0.022	7.77	6	0.255
Depressive symptoms state										
1c. Free loadings	39.60	30	1.048	0.998	0.996	0.024	0.019			
2c. Constraints on loadings	48.92	36	1.014	0.997	0.995	0.026	0.025	9.60	6	0.142
3c. Constraints on loadings and intercepts	66.18	42	1.014	0.995	0.992	0.033	0.028	17.26	6	0.008
Depressive symptoms trait										
1d. Free loadings	44.90	30	1.142	0.997	0.994	0.030	0.015			
2d. Constraints on loadings	49.08	36	1.124	0.998	0.996	0.026	0.016	3.76	6	0.709
3d. Constraints on loadings and intercepts	80.23	42	1.102	0.993	0.989	0.041	0.019	34.28	6	0.000

Note. N = 1,087. Results based on the MLR estimator (Muthén, & Muthén, 1998-2017). SF = Scaling Factor; CFI = comparative fit index; TLI = Tucker Lewis Index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

Supplemental Material B2: Bivariate Cross-Sectional Latent Correlations between Self-Esteem and Depressive Symptoms

Table B5
Bivariate Cross-Sectional Latent Correlations between Self-Esteem and Depressive Symptoms

	Exploratory study				Confirmatory study			
	A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state	A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state
r (SE1, D1)	-.75	-.82	-.85	-.77	-.71	-.83	-.78	-.84
r (SE2, D2)	-.70	-.85	-.95	-.90	-.70	-.86	-.91	-.87
r (SE3, D3)	-.81	-.85	-.91	-.90	-.68	-.82	-.88	-.91
r (SE4, D4)	-.69	-.82	-.86	-.88	-.72	-.85	-.86	-.86

Note. Exploratory study: N_{Total} = 683; N_A = 179; N_B = 163; N_C = 170; N_D = 171. Confirmatory study: N_{Total} = 1087; N_A = 280; N_B = 269; N_C = 273; N_D = 265.

Supplemental Material B3: Latent Descriptive Statistics, McDonalds ω, and Raw Rank-Order Stabilities for Self-Esteem and Depressive Symptoms Separated by Gender

Table B6
Latent Descriptive Statistics, McDonalds ω, and *Raw Rank-Order Stabilities for Self-Esteem Separated by Gender*

		Exploratory study						Confirmatory study					
		Female			Male			Female			Male		
		M (SD)	ω	$r_{t,t-1}$	*M (SD)*	ω	$r_{t,t-1}$	*M (SD)*	ω	$r_{t,t-1}$	*M (SD)*	ω	$r_{t,t-1}$
T1	State	2.94 (0.60)	.90		3.07 (0.60)	.88		2.94 (0.60)	.92		3.07 (0.57)	.91	
	Trait	3.01 (0.57)	.92		3.19 (0.53)	.91		3.01 (0.48)	.91		3.19 (0.53)	.90	
T2	State	3.00 (0.57)	.91	.71	3.13 (0.50)	.87	.90	3.00 (0.57)	.92	.82	3.13 (0.50)	.90	.81
	Trait	3.01 (0.54)	.93	.95	3.21 (0.52)	.89	.85	3.01 (0.54)	.91	.92	3.21 (0.52)	.90	.96
T3	State	3.00 (0.59)	.92	.77	3.16 (0.53)	.92	.90	3.00 (0.59)	.91	.81	3.16 (0.53)	.92	.80
	Trait	3.07 (0.56)	.92	.96	3.23 (0.53)	.88	.96	3.07 (0.56)	.92	.91	3.23 (0.53)	.92	.95
T4	State	3.02 (0.60)	.92	.72	3.13 (0.55)	.94	.81	3.02 (0.60)	.92	.85	3.13 (0.56)	.91	.78
	Trait	3.10 (0.58)	.94	.93	3.24 (0.58)	.88	.90	3.10 (0.57)	.92	.92	3.24 (0.58)	.93	.94

Note. Exploratory study: $N_{State,\ female}$ = 238, $N_{Trait,\ female}$ = 257, $N_{State,\ male}$ = 84; $N_{Trait,\ male}$ = 69; Confirmatory study: $N_{State,\ female}$ = 355, $N_{Trait,\ female}$ = 369, $N_{State,\ male}$ = 146 ; $N_{Trait,\ male}$ = 133; Includes only individuals who participated at T1 because information about gender is only available at T1.

Table B7
Latent Descriptive Statistics, McDonalds ω, and *Raw Rank-Order Stabilities for Depressive Symptoms Separated by Gender*

		Exploratory study						Confirmatory study					
		Female			Male			Female			Male		
		M (SD)	ω	$r_{t,t-1}$	*M (SD)*	ω	$r_{t,t-1}$	*M (SD)*	ω	$r_{t,t-1}$	*M (SD)*	ω	$r_{t,t-1}$
T1	State	1.02 (0.52)	.90		0.88 (0.48)	.89		1.09 (0.51)	.89		0.95 (0.51)	.89	
	Trait	0.91 (0.51)	.91		0.84 (0.39)	.85		1.03 (0.54)	.91		0.96 (0.48)	.89	
T2	State	1.06 (0.56)	.92	.69	0.92 (0.52)	.91	.65	1.10 (0.51)	.90	.72	0.97 (0.52)	.91	.71
	Trait	0.95 (0.53)	.92	.87	0.87 (0.47)	.89	.94	1.04 (0.57)	.93	.89	0.99 (0.52)	.91	.91
T3	State	1.08 (0.56)	.93	.63	0.97 (0.63)	.92	.78	1.09 (0.54)	.91	.73	1.00 (0.55)	.91	.80
	Trait	1.00 (0.56)	.92	.91	0.91 (0.49)	.91	.93	1.04 (0.56)	.93	.88	0.95 (0.53)	.92	.86
T4	State	1.20 (0.55)	.91	.75	1.14 (0.62)	.93	.71	1.14 (0.54)	.91	.71	1.03 (0.56)	.91	.76
	Trait	0.97 (0.53)	.91	.89	0.83 (0.41)	.89	.93	1.02 (0.58)	.93	.86	0.93 (0.54)	.93	.89

Note. Exploratory study: $N_{State,\ female}$ = 258, $N_{Trait,\ female}$ = 237, $N_{State,\ male}$ = 77; $N_{Trait,\ male}$ = 76; Confirmatory study: $N_{State,\ female}$ = 362, $N_{Trait,\ female}$ = 362, $N_{State,\ male}$ = 143; $N_{Trait,\ male}$ = 136; Includes only individuals who participated at T1 because information about gender is only available at T1.

Table B8

Bivariate Cross-Sectional Latent Correlations between Self-Esteem and Depressive Symptoms Separated by Gender

	Exploratory study (female/male)				Confirmatory study (female/male)			
	A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state	A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state
r (SE1, D1)	-.72/-.75[a]	-.84/-.64[a]	-.86/-.70[a]	-.75/-.68[a]	-.72/-.71	-.82/-.85	-.79/-.79	-.81/-.91
r (SE2, D2)	-.73/-.47[a]	-.88/-.71[a]	-.94/-.86[a]	-.90/-.74[a]	-.71/-.75	-.85/-.91	-.90/-.90	-.87/-.83
r (SE3, D3)	-.83/-.61[a]	-.87/-.72[a]	-.91/-.80[a]	-.89/-.88[a]	-.66/-.74	-.80/-.87	-.90/-.85	-.92/-.89
r (SE4, D4)	-.66/-.72[a]	-.85/-.69[a]	-.86/-.76[a]	-.89/-.81[a]	-.73/-.69	-.84/-.86	-.86/-.87	-.85/-.92

Note. Sample sizes (female/male) in the exploratory study: N_{Total} = 495/153; N_A = 135/37; N_B = 122/32; N_C = 115/44; N_D = 123/40; Sample sizes (female/male) in the confirmatory study: N_{Total} = 724/279; N_A = 182/71; N_B = 187/62; N_C = 175/74; N_D = 180/72.

[a] Due to the small sample size for males in the exploratory study, we were not able to run latent analyses. This is why these are manifest correlations.

Supplemental Material B4: (Latent State-Trait) Cross-Lagged Panel Models without Longitudinal Constraints on the Regression Coefficients

Table B9 and B10 depict the results based on the (latent state-trait) cross-lagged panel models without longitudinal constraints on the regression coefficients. Results based on models without longitudinal constraints on all regression coefficients deviated slightly from those with longitudinal constraints. In both the CLPM and the LST-CLPM, in both studies, the parameters were not statistically significant at all time points. Additionally, in the CLPM in the exploratory study in Condition C, the path from depressive symptoms to self-esteem from T1 to T2 was statistically significant, whereas in Condition D, none of the single paths were statistically significant. In the LST-CLPM, in the exploratory study in Condition D, the path from self-esteem to depressive symptoms was statistically significant, whereas in the confirmatory study in Condition C, none of the cross-lagged paths were statistically significant, and in Condition D, the path from depressive symptoms to self-esteem from T3 to T4 was statistically significant.

Table B9

Estimates [and 95% Confidence Intervals] of the Structural Coefficients from the Cross-Lagged Panel Model (CLPM) without Longitudinal Constraints on the Regression Coefficients

Coefficient		Exploratory study				Confirmatory study			
		A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state	A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state
		β [CI]	β [CI]	β [CI]	β [CI]	β [CI]	β [CI]	β [CI]	β [CI]
Stability									
Self-esteem	T1→T2	**1.03 [.91, 1.14]**	**.98 [.80, 1.17]**	**.52 [.21, .82]**	**.55 [.30, .81]**	**.97 [.89, 1.06]**	**.90 [.77, 1.02]**	**.62 [.42, .81]**	**.94 [.74, 1.15]**
	T2→T3	**.95 [.86, 1.04]**	**.96 [.77, 1.15]**	.77 [-.19, 1.72]	.67 [-.12, 1.46]	**1.00 [.90, 1.09]**	**1.07 [.91, 1.24]**	**.46 [.05, .87]**	**.78 [.50, 1.06]**
	T3→T4	**.99 [.87, 1.11]**	**.93 [.75, 1.11]**	**.75 [.20, 1.31]**	.39 [-.03, .82]	**1.02 [.93, 1.10]**	**.97 [.85, 1.09]**	**.82 [.57, 1.07]**	**.87 [.48, 1.26]**
Depressive symptoms	T1→T2	**.43 [.20, .66]**	**.70 [.50, .90]**	**.68 [.40, .97]**	**.55 [.30, .80]**	**.49 [.31, .66]**	**.76 [.61, .90]**	**.68 [.52, .83]**	**.59 [.37, .82]**
	T2→T3	**.37 [.14, .60]**	**.92 [.71, 1.13]**	**1.15 [.21, 2.09]**	.47 [-.19, 1.13]	**.62 [.44, .80]**	**.87 [.70, 1.04]**	**1.07 [.65, 1.49]**	**.72 [.47, .98]**
	T3→T4	**.72 [.49, .95]**	**.68 [.46, .89]**	**.81 [.43, 1.18]**	**.74 [.35, 1.13]**	**.54 [.38, .70]**	**.76 [.61, .91]**	**.70 [.38, 1.01]**	**.81 [.41, 1.21]**
Cross-lagged									
	SE1→D2	**-.35 [-.56, -.14]**	**-.25 [-.46, -.03]**	-.23 [-.54, .07]	-.20 [-.48, .09]	**-.26 [-.42, -.09]**	*-.16* [-.32, .00]	**-.28 [-.44, -.12]**	*-.22* [-.45, .03]
	SE2→D3	**-.46 [-.67, -.24]**	.01 [-.23, .24]	.22 [-.75, 1.19]	-.21 [-.88, .45]	*-.17 [-.35, .01]*	-.04 [-.22, .14]	.22 [-.21, .66]	-.07 [-.35, .21]
	SE3→D4	-.08 [-.35, .19]	**-.24 [-.47, -.01]**	-.18 [-.50, .27]	.03 [-.38, .44]	**-.28 [-.43, -.12]**	**-.16 [-.32, -.00]**	-.17 [-.51, .16]	.05 [-.36, .46]
	D1→SE2	.11 [-.04, .26]	.04 [-.17, .26]	**-.36 [-.66, -.05]**	-.22 [-.48, .04]	.04 [-.06, .15]	-.05 [-.19, .09]	**-.26 [-.46, -.06]**	.12 [-.11, .34]
	D2→SE3	-.01 [-.13, .10]	-.02 [-.23, .19]	-.11 [-1.07, .85]	-.12 [-.92, .68]	.06 [-.07, .20]	*.17* [-.02, .36]	*-.38* [-.79, .03]	-.06 [-.35, .23]
	D3→SE4	.07 [-.08, .21]	.03 [-.17, .24]	-.09 [-.67, .50]	-.36 [-.75, .03]	**.12 [.00, .23]**	.04 [-.10, .18]	-.01 [-.29, .27]	.04 [-.37, .44]

Note. Exploratory study: N_{Total} = 683; N_A = 179; N_B = 163; N_C = 170; N_D = 171. Confirmatory study: N_{Total} = 1087; N_A = 280; N_B = 269; N_C = 273; N_D = 265. Condition A: Self-Esteem trait, Depressive symptoms state; Condition B: Self-Esteem trait, Depressive symptoms trait; Condition C: Self-Esteem state, Depressive symptoms trait; Condition D: Self-esteem state, Depressive symptoms state. Bold indicates that the 95% confidence interval did not include zero. Italics indicate that the 90% confidence interval did not include zero. All coefficients are standardized.

Table B10

Estimates [and 95% Confidence Intervals] of the Structural Coefficients from the Latent State-Trait Cross-Lagged Panel Model (LST-CLPM) without Longitudinal Constraints on the Regression Coefficients

Coefficient		Exploratory study				Confirmatory study			
		A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state	A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state
		β^a [CI]	β^a [CI]	β^a [CI]	β^a [CI]	β^a [CI]	β^a [CI]	β^a [CI]	β^a [CI]
Stability									
Self-esteem	T1→T2	.00 [-.18, .18]	-.05 [-.24, .14]	.28 [-.27, .84]	-.20 [-.54, .13]	-.06 [-.19, .08]	-.05 [-.24, .13]	.40 [-.93, 1.72]	.06 [-.23, .35]
	T2→T3	.02 [-.15, .18]	-.07 [-.30, .16]	1.39 [-.82, 3.59]	-.30 [-.63, .03]	-.04 [-.15, .06]	.10 [-.09. .28]	.58 [-1.72, 2.89]	-.17 [-.46, .12]
	T3→T4	-.10 [-.32, .12]	-.12 [-.37, .13]	.55 [-.01, 1.10]	**-.75** **[-1.33, -.18]**	-.01 [-.12, .14]	.12 [-.04, .28]	.60 [-2.06, 3.25]	**-.31** **[-.69, -.07]**
Depressive symptoms	T1→T2	.46 [-.28, 1.20]	.21 [-.55, .97]	-.11 [-.42, .19]	.28 [-.08, .63]	**.27** **[.07, .48]**	.19 [-.01, .40]	-.09 [-.32, .14]	.20 [-.07, .47]
	T2→T3	.32 [-.30, .94]	.23 [-.38, .84]	-1.04 [-2.78, .70]	.27 [-.16, .69]	**.36** **[.14, .58]**	.15 [-.13, .42]	-.12 [-.62, .38]	.13 [-.17, .42]
	T3→T4	.79 [-.07, 1.64]	.10 [-.65, .85]	-.20 [-.61, .20]	**.84** **[.35, 1.33]**	**.31** **[.10, .52]**	.08 [-.10, .27]	-.16 [-.56, .23]	*.28* [-.08, .64]
Cross-lagged									
	SE1→D2	.22 [-.27, .70]	.02 [-.60, .64]	-.27 [-.63, .09]	.12 [-.20, .45]	**.23** **[.02, .04]**	.06 [-.13, .25]	-.16 [-.40, .07]	-.05 [-.33, .22]
	SE2→D3	.06 [-.39, .50]	.11 [-.43, .65]	-1.27 [-3.02, .47]	.12 [-.29, .53]	**.26** **[.04, .47]**	.06 [-.19, .30]	-.19 [-.73, .35]	-.04 [-.31, .23]
	SE3→D4	.56 [-.12, 1.24]	.02 [-.63, .66]	-.39 [-.82, .04]	**.70** **[.21, 1.93]**	*.17* [-.03, .37]	-.07 [-.24, .10]	-.22 [-.60, .17]	.11 [-.26, -48]
	D1→SE2	-.08 [-.30, .13]	-.11 [-.29, .08]	.13 [-.28, .52]	-.24 [-.54, .06]	-.04 [-.14, .06]	-.07 [-.13, .26]	.36 [-.75, 1.46]	-.08 [-.35, .20]
	D2→SE3	-.04 [-.18, .10]	-.07 [-.35, .10]	1.17 [-1.00, 3.35]	**-.45** **[-.77, -.13]**	-.06 [-.17, .04]	.07 [-.13, .26]	.50 [-1.63, 2.64]	*-.27* [-.57, .04]
	D3→SE4	-.19 [-.41, .03]	-.13 [-.43, .08]	0.39 [-.13, .91]	**-.86** **[-1.34, -.35]**	-.02 [-.13, .10]	.08 [-.08, .24]	.53 [-1.76, 2.82]	**-.49** [-.86, -.12]

Note. Exploratory study: N_{Total} = 683; N_A = 179; N_B = 163; N_C = 170; N_D = 171. Confirmatory study: N_{Total} = 1087; N_A = 280; N_B = 269; N_C = 273; N_D = 265. Condition A: Self-esteem trait, Depressive symptoms state; Condition B: Self-esteem trait, Depressive symptoms trait; Condition C: Self-esteem state, Depressive symptoms trait; Condition D: Self-esteem state, Depressive symptoms state. In the confirmatory study, in this specific model, we did

not include gender as an auxiliary variable because the model had convergence issues when gender was included. Bold indicates that the 95% confidence interval did not include zero. Italics indicate that the 90% confidence interval did not include zero. All coefficients are standardized.
[a] Refers to the meaning of the structural coefficients in the LST-CLPM.

Supplemental Material B5: (Latent State-Trait) Cross-Lagged Panel Models for Females

Results on the cross-lagged panel model based on the female samples in the exploratory and confirmatory studies are presented in Table B11 (with longitudinal constraints) and Table B12 (without longitudinal constraints). All results were largely unchanged compared with the total samples.

Results on the latent state-trait cross-lagged panel model based on the female sample in the confirmatory study are presented in Table B13. Unfortunately, it was not possible to run these analyses in the exploratory study because the sample size per condition was too small. The results in the confirmatory study were largely unchanged in comparison with the total sample.

Table B11

Estimates [and 95% Confidence Intervals] of the Structural Coefficients from the Cross-Lagged Panel Model (CLPM) for Females

Coefficient		Exploratory study				Confirmatory study			
		A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state	A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state
		β [CI]	β [CI]	β [CI]	β [CI]	β [CI]	β [CI]	β [CI]	β [CI]
Stability									
Self-esteem	T1→T2	**.99 [.93; 1.06]**	**1.01 [.92; 1.09]**	**.61 [.17; 1.04]**	**.41 [.14; .68]**	**1.01 [.95; 1.07]**	**.98 [.92; 1.05]**	**.67 [.47; .86]**	**.86 [.71; 1.01]**
	T2→T3	**.99 [.93; 1.05]**	**1.04 [.96; 1.12]**	**.59 [.11; 1.08]**	**.43 [.16; .70]**	**.99 [.93; 1.06]**	**.97 [.90; 1.04]**	**.64 [.45; .82]**	**.87 [.70; 1.03]**
	T3→T4	**.96 [.89; 1.03]**	**.97 [.89; 1.05]**	**.57 [.13; 1.02]**	**.40 [.15; .66]**	**.99 [.92; 1.05]**	**.97 [.89; 1.05]**	**.64 [.43; .84]**	**.87 [.71; 1.02]**
Depressive symptoms	T1→T2	**.44 [.24; .64]**	**.75 [.63; .88]**	**.75 [.54; .97]**	**.56 [.33; .78]**	**.52 [.36; .68]**	**.78 [.68; .88]**	**.71 [.55; .87]**	**.62 [.44; .80]**
	T2→T3	**.42 [.26; .58]**	**.75 [.62; .88]**	**.79 [.56; 1.01]**	**.58 [.34; .82]**	**.49 [.34; .65]**	**.82 [.73; .90]**	**.68 [.50; .85]**	**.63 [.43; .83]**
	T3→T4	**.45 [.25; .65]**	**.73 [.60; .85]**	**.79 [.57; 1.01]**	**.61 [.36; .87]**	**.54 [.38; .69]**	**.79 [.69; .89]**	**.70 [.56; .83]**	**.61 [.42; .80]**
Cross-lagged									
	SE1→D2	**-.39 [-.58; -.20]**	**-.18 [-.30; -.07]**	-.14 [-.36; .08]	-.09 [-.32; .15]	**-.26 [-.40; -.12]**	**-.11 [-.20; -.02]**	**-.23 [-.39; -.07]**	*-.16 [-.34; .02]*
	SE2→D3	**-.37 [-.56; -.17]**	**-.19 [-.30; -.07]**	-.14 [-.35; .08]	-.09 [-.32; .15]	**-.24 [-.37; -.10]**	**-.11 [-.20; -.02]**	**-.21 [-.34; -.07]**	*-.16 [-.34; .02]*
	SE3→D4	**-.37 [-.56; -.18]**	**-.18 [-.29; -.06]**	-.14 [-.35; .08]	-.09 [-.33; .15]	**-.24 [-.38; -.10]**	**-.11 [-.20; -.02]**	**-.20 [-.34; -.06]**	*-.16 [-.33; .02]*
	D1→SE2	.04 [-.03; .12]	.06 [-.02; .15]	-.26 [-.71; .20]	**-.30 [-.52; -.08]**	**.08 [.01; .15]**	*.06 [-.01; .12]*	**-.22 [-.40; -.03]**	.02 [-.13; .17]
	D2→SE3	.04 [-.03; .12]	.07 [-.02; .15]	-.27 [-.73; .20]	**-.33 [-.57; -.09]**	**.08 [.01; .15]**	*.06 [-.01; .13]*	**-.22 [-.42; -.03]**	.02 [-.13; .17]
	D3→SE4	.04 [-.03; .12]	.06 [-.02; .15]	-.26 [-.71; .20]	**-.32 [-.56; -.08]**	**.08 [.01; .16]**	*.06 [-.01; .13]*	**-.24 [-.43; -.04]**	.02 [-.13; .17]

Note. Female sample sizes in the exploratory study: N_{Total} = 495; N_A = 135; N_B = 122; N_C = 115; N_D = 123; Female sample sizes in the confirmatory study: N_{Total} = 724; N_A = 182; N_B = 187; N_C = 175; N_D = 180. Condition A: Self-esteem trait, Depressive symptoms state; Condition B: Self-esteem trait, Depressive symptoms trait; Condition C: Self-esteem state, Depressive symptoms trait; Condition D: Self-esteem state, Depressive symptoms state. Bold indicates that the 95% confidence interval did not include zero. Italics indicate that the 90% confidence interval did not include zero. All coefficients are standardized.

Table B12

Estimates [and 95% Confidence Intervals] of the Structural Coefficients from the Cross-Lagged Panel Model (CLPM) for Females without Longitudinal Constraints on the Regression Coefficients

Coefficient		Exploratory study				Confirmatory study			
		A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state	A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state
		β [CI]	β [CI]	β [CI]	β [CI]	β [CI]	β [CI]	β [CI]	β [CI]
Stability									
	T1→T2	**1.01** **[.90; 1.11]**	**1.03** **[.82; 1.25]**	**.50** **[.15; .85]**	**.44** **[.15; .73]**	**.91** **[.80; 1.02]**	**.88** **[.75; 1.01]**	**.68** **[.43; .92]**	**.92** **[.72; 1.13]**
Self-esteem	T2→T3	**1.00** **[.91; 1.10]**	**.99** **[.71; 1.26]**	*1.07* *[-.03; 2.16]*	.48 [-.56; 1.53]	**1.03** **[.90; 1.17]**	**1.05** **[.86; 1.25]**	.32 [-.12; .75]	**.84** **[.52; 1.16]**
	T3→T4	**1.02** **[.88; 1.16]**	**.97** **[.77; 1.17]**	**.88** **[.15; 1.60]**	.26 [-.28; .80]	**1.02** **[.92; 1.11]**	**.99** **[.86; 1.12]**	**.85** **[.63; 1.08]**	**.76** **[.18; 1.34]**
	T1→T2	**.43** **[.21; .65]**	**.58** **[.37; .80]**	**.73** **[.44; 1.02]**	**.57** **[.27; .86]**	**.50** **[.28; .73]**	**.76** **[.58; .94]**	**.62** **[.44; .81]**	**.59** **[.36; .83]**
Depressive symptoms	T2→T3	.18 [-.09; .46]	**1.09** **[.78; 1.39]**	.80 [-.15; 1.75]	.53 [-.37; 1.43]	**.55** **[.31; .79]**	**.92** **[.75; 1.09]**	**1.19** **[.87; 1.51]**	**.60** **[.29; .90]**
	T3→T4	**.85** **[.58; 1.12]**	**.59** **[.35; .82]**	**.90** **[.14; 1.66]**	**.71** **[.22; 1.19]**	**.47** **[.28; .67]**	**.70** **[.54; .87]**	**.64** **[.27; 1.00]**	**.89** **[.28; 1.49]**
Cross-lagged									
	SE1→D2	**-.41** **[-.61; -.21]**	**-.39** **[-.61; -.16]**	-.16 [-.48; .16]	-.13 [-.47; .21]	**-.24** **[-.45; -.03]**	-.14 [-.34; .05]	**-.34** **[-.52; -.16]**	-.21 [-.46; .05]
	SE2→D3	**-.62** **[-.87; -.36]**	.17 [-.16; .51]	-.14 [-1.12; .84]	-.10 [-1.02; .81]	-.18 [-.43; .05]	.01 [-.18; .20]	*.34* *[-.02; .71]*	-.21 [-.53; .11]
	SE3→D4	.07 [-.25; .40]	**-.33** **[-.56; -.09]**	-.01 [-.80; .77]	.01 [-.50; .52]	**-.33** **[-.52; -.14]**	**-.22** **[-.40; -.05]**	-.26 [-.64; .12]	.16 [-.47; .78]
	D1→SE2	.06 [-.07; .20]	.10 [-.15; .35]	**-.38** **[-.74; -.01]**	*-.30* *[-.62; .01]*	-.04 [-.17; .10]	-.06 [-.21; .10]	-.18 [-.42; .06]	.10 [-.13; .34]
	D2→SE3	.06 [-.06; .18]	.00 [-.29; .30]	.19 [-.92; 1.30]	-.26 [-1.32; .79]	.12 [-.07; .31]	.16 [-.06; .38]	**-.55** **[-.98; -.12]**	.00 [-.34; .34]
	D3→SE4	.10 [-.06; .26]	.06 [-.18; .29]	.05 [-.72; .82]	*-.46* *[-.94; .02]*	*.13* *[.00; .26]*	.08 [-.07; .24]	-.01 [-.26; .23]	-.10 [-.68; .49]

Note. Female sample sizes in the exploratory study: N_{Total} = 495; N_A = 135; N_B = 122; N_C = 115; N_D = 123; Female sample sizes in the confirmatory study: N_{Total} = 724; N_A = 182; N_B = 187; N_C = 175; N_D = 180. Condition A: Self-esteem trait, Depressive symptoms state; Condition B: Self-esteem trait, Depressive symptoms trait; Condition C: Self-esteem state, Depressive symptoms trait; Condition D: Self-esteem state, Depressive symptoms state. Bold indicates that the 95% confidence interval did not include zero. Italics indicate that the 90% confidence interval did not include zero. All coefficients are standardized.

Table B13

Estimates [and 95% Confidence Intervals] of the Structural Coefficients from the Latent State-Trait Cross-Lagged Panel Model (LST-CLPM) for Females in the Confirmatory Study

Coefficient		Confirmatory study: Model with longitudinal constraints				Confirmatory study: Model without longitudinal constraints			
		A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state	A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state
		β^a [CI]	β^a [CI]	β^a [CI]	β^a [CI]	β^a [CI]	β^a [CI]	β^a [CI]	β^a [CI]
Stability									
Self-esteem	T1→T2	-.07 [-.21; .06]	.07 [-.09; .24]	*.28* *[-.04; .61]*	-.03 [-.29; .23]	-.09 [-.26; .08]	-.06 [-.24; .12]	.29 [-.21; .79]	.04 [-.25; .34]
	T2→T3	-.07 [-.19; .06]	.08 [-.10; .25]	*.31* *[-.04; .66]*	-.03 [-.28; .21]	-.06 [-.19; .07]	.09 [-.10; .28]	.39 [-.26; 1.04]	-.16 [-.54; .21]
	T3→T4	-.07 [-.19; .06]	.08 [-.09; .24]	*.27* *[-.06; .60]*	-.03 [-.29; .23]	-.04 [-.19; .10]	*.17* *[.02; .33]*	.42 [-.57; 1.41]	**-.45** **[-.93; .04]**
Depressive symptoms	T1→T2	**.34** **[.12; .57]**	.07 [-.09; .23]	-.10 [-.30; .10]	.13 [-.17; .44]	**.30** **[.03; .57]**	.14 [-.06; .34]	-.08 [-.35; .18]	.23 [-.08; .54]
	T2→T3	**.30** **[.09; .51]**	.07 [-.10; .25]	-.11 [-.32; .11]	.13 [-.16; .42]	**.38** **[.10; .66]**	.12 [-.19; .43]	-.13 [-.50; .24]	.01 [-.38; .39]
	T3→T4	**.34** **[.12; .56]**	.07 [-.10; .24]	-.11 [-.32; .10]	.13 [-.17; .44]	**.34** **[.09; .59]**	.02 [-.15; .20]	-.23 [-.58; .11]	.29 [-.17; .75]
Cross-lagged									
	SE1→D2	**.30** **[.07; .53]**	-.05 [-.21; .11]	-.19 [-.41; .03]	-.10 [-.40; .20]	**.30** **[.04; .56]**	.01 [-.18; .20]	-.15 [-.41; .12]	-.05 [-.37; .27]
	SE2→D3	**.26** **[.06; .46]**	-.05 [-.22; .11]	-.19 [-.42; .04]	-.09 [-.38; .19]	**.33** **[.06; .61]**	.04 [-.24; .32]	-.20 [-.61; .20]	-.18 [-.52; .15]
	SE3→D4	**.27** **[.07; .47]**	-.05 [-.21; .11]	-.17 [-.39; .04]	-.10 [-.40; .20]	*.23* *[-.01; .47]*	*-.14* *[-.30; .02]*	-.30 [-.66; .06]	.12 [-.35; .59]
	D1→SE2	-.02 [-.12; .07]	.05 [-.10; .19]	.21 [-.06; .48]	*-.22* *[-.47; .04]*	-.01 [-.15; .13]	-.07 [-.23; .08]	.28 [-.18; .75]	-.14 [-.43; .16]
	D2→SE3	-.02 [-.11; .07]	.05 [-.11; .21]	.23 [-.06; .53]	*-.21* *[-.45; .03]*	-.05 [-.18; .07]	.07 [-.13; .27]	.28 [-.35; .91]	-.28 [-.68; .12]
	D3→SE4	-.02 [-.12; .07]	.05 [-.11; .21]	.22 [-.06; .51]	*-.22* *[-.49; .04]*	-.02 [-.15; .12]	.13 [-.03; .29]	.38 [-.50; 1.26]	**-.66** **[-1.15; -.16]**

Note. Female sample sizes in the confirmatory study: N_{Total} = 724; N_A = 182; N_B = 187; N_C = 175; N_D = 180. Condition A: Self-esteem trait, Depressive symptoms state; Condition B: Self-esteem trait, Depressive symptoms trait; Condition C: Self-esteem state, Depressive symptoms trait; Condition D:

Self-esteem state, Depressive symptoms state. Bold indicates that the 95% confidence interval did not include zero. Italics indicate that the 90% confidence interval did not include zero. All coefficients are standardized.

[a] Refers to the meaning of the structural coefficients in the LST-CLPM.

Supplemental Material B6: Random Intercept Cross-Lagged Panels Models

As proposed by Hamaker et al. (2015), the Random Intercept Cross-Lagged Panel Model (RI-CLPM) aims at analyzing reciprocal relations between constructs over time while controlling for stable trait differences in these constructs. The RI-CLPM and the Latent State-Trait Cross-Lagged Panel Model (LST-CLPM) are comparable models regarding their structure. However, they differ on particular important attributes, which we describe in further detail below (for a discussion on the univariate, autoregressive versions of these models, see Cole, Martin, & Steiger, 2005, and Prenoveau, 2016; see also Steyer et al., 2015).

In the RI-CLPM, the predictors in each autoregressive and cross-lagged path are structured state residuals that represent the time-point-specific within-person deviations from the random intercept factor (trait factor). Thus, the cross-lagged coefficients indicate how between-person differences in the within-person deviations in one variable are associated with between-person differences in the within-person deviations in the other variable at the next time point (while controlling for previous between-person differences in the within-person deviations in the latter variable). Therefore, estimating the cross-lagged effects based on the state residuals (i.e., within-person deviations) as proposed within the RI-CLPM is independent of the absolute trait level of the variables.

By contrast, in the LST-CLPM (see also AR CL model with unit effects in Zyphur, Allison, et al., 2019), the predictors in each autoregressive and cross-lagged path are not the time-point-specific deviations, but the latent state variables themselves. Latent state variables represent the (measurement free) response behaviors at specific time points. The interpretation of the respective outcome variables is comparable to the RI-CLPM because the variance that can be explained is the variance that is not explained by the trait factor (i.e., in this regard it is the state residual). However, in contrast to the RI-CLPM, the outcome variables in the LST-CLPM are directly predicted by the trait factor of the respective constructs. Thus, the autoregressive and cross-lagged coefficients are directly conditioned on the trait level of the respective outcome variable. Consequently, cross-lagged and autoregressive coefficients indicate how between-persons differences in one variable are associated with between-person differences in the other variable while controlling for the trait level and previous between-person differences in the latter variable. We think that this model is best for capturing our idea of combining latent state-trait analyses and cross-lagged panel models by explicitly controlling for the trait variance when estimating the cross-lagged coefficients rather than estimating the

cross-lagged coefficients based on residuals that remained unexplained by the trait factor. However, we think that both models are reasonable, but they represent slightly different theoretical ideas.

Table B14 depicts the standardized regression coefficients and 95% confidence intervals for the RI-CLPM in the exploratory and the confirmatory study. We compared the results of the RI-CLPM with the results of the LST-CLPM (see Table 4 in the main text) in the different time frame conditions in the exploratory and the confirmatory study. Overall, we made four major observations: First, in line with the findings from the LST-CLPM, the results of the RI-CLPM were not as consistent across the two studies as the results of the regular CLPM. Second, across both studies and across Conditions A, B, and C, there were some differences between the results from LST-CLPM and the RI-CLPM. Third, however, in both studies the overall patterns of the LST-CLPM and the RI-CLPM were very similar regarding Condition D (self-esteem state, depressive symptoms state). Fourth, across both studies, the confidence intervals in the RI-CLPM were typically—in some cases even dramatically—larger than the confidence intervals in the LST-CLPM.

In sum, the two models produced some different results when at least one construct was measured as a trait. However, the patterns of results were very similar when both constructs were measured as states. A possible explanation could be that, when measured as states, latent state variables and the respective latent state residuals are more similar than in the other conditions because there is a higher amount of state residual variance in state measures than in trait measures (see, e.g., Figure 2 in the main text). Thus, controlling for the trait level in the outcome variable is less consequential for the partial regression coefficients in the LST-CLPM when the trait level explains less variance in the outcome variable (as it is typically the case when using state measures). This circumstance does not only hold for the similarity between the LST-CLPM and the RI-CLPM but also for their comparisons with the regular CLPM. Controlling for the trait variance (as implemented in both the LST-CLPM and the RI-CLPM) is less influential when there is less trait variance in the constructs. Hence, the results in Condition D are very similar in all three models, that is, in the CLPM (see Table 3), the LST-CLPM (see Table 4) and the RI-CLPM (see Table B14).

Table B14

Estimates [and 95% Confidence Intervals] of the Structural Coefficients from the Random Intercept Cross-Lagged Panel Model (RI-CLPM)

Coefficient		Exploratory study				Confirmatory study			
		A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state	A S trait, D state	B S trait, D trait	C S state, D trait	D S state, D state
		β [CI]	β [CI]	β [CI]	β [CI]	β [CI]	β [CI]	β [CI]	β [CI]
Stability									
Self-esteem	T1→T2	**.57** **[.12; 1.03]**	.22 [-.36; .80]	.35 [-.01; .72]	-.14 [-.55; .28]	.17 [-.51; .86]	**.55** **[.22; .87]**	**.49** **[.21; .78]**	.63 [-.66; 1.93]
	T2→T3	**.61** **[.28; .93]**	.44 [-.70; 1.57]	.28 [-.06; .63]	-.23 [-.97; .51]	.08 [-.18; .34]	**.57** **[.25; .88]**	**.46** **[.18; .74]**	.52 [-.92; 1.96]
	T3→T4	**.39** **[-.21; 1.00]**	.08 [-.22; .39]	.30 [-.03; .64]	-.13 [-.51; .24]	.10 [-.29; .49]	**.63** **[.26; .99]**	**.42** **[.11; .73]**	.55 [-.89; 2.00]
Depressive symptoms	T1→T2	.30 [-.11; .71]	.49 [-.14; 1.13]	**.17** **[-.12; .45]**	**.38** **[.14; .62]**	**.39** **[.09; .69]**	.14 [-.19; .47]	.06 [-.11; .24]	.12 [-.90; 1.13]
	T2→T3	.33 [-.06; .71]	.39 [-.23; 1.02]	**.37** **[.06; .69]**	**.52** **[.22; .83]**	**.30** **[-.04; .64]**	.20 [-.28; .68]	.08 [-.13; .29]	.10 [-.77; .98]
	T3→T4	.34 [-.11; .78]	.36 [-.20; .92]	**.39** **[.05; .73]**	**.49** **[.17; .81]**	**.35** **[.01; .69]**	.18 [-.26; .62]	.09 [-.16; .35]	.10 [-.78; .99]
Cross-lagged									
	SE1→D2	-.27 [-1.02; .47]	.02 [-.59; .64]	**-.35** **[-.66; -.04]**	.14 [-.15; .43]	.10 [-.24; .43]	**-.33** **[-.71; .04]**	**-.40** **[-.67; -.13]**	-.24 [-1.50; 1.02]
	SE2→D3	-.24 [-.95; .47]	.02 [-.57; .61]	**-.32** **[-.64; .00]**	.20 [-.23; .62]	.05 [-.13; .23]	**-.38** **[-.78; .02]**	**-.32** **[-.57; -.06]**	-.19 [-1.27; .89]
	SE3→D4	-.20 [-.92; .52]	.01 [-.25; .27]	**-.38** **[-.72; -.03]**	.15 [-.14; .44]	.08 [-.18; .34]	**-.38** **[-.77; .01]**	**-.31** **[-.57; -.04]**	-.20 [-1.29; .89]
	D1→SE2	-.15 [-.39; .09]	-.05 [-.49; .40]	**-.21** **[-.56; .14]**	**-.39** **[-.69; -.09]**	.00 [-.37; .36]	.07 [-.16; .30]	-.03 [-.18; .12]	.17 [-.76; 1.10]
	D2→SE3	-.20 [-.52; .12]	-.08 [-.79; .63]	**-.41** **[-.70; -.11]**	**-.65** **[-1.18; -.11]**	.00 [-.23; .23]	.09 [-.21; .39]	-.05 [-.27; .18]	.15 [-.70; 1.00]
	D3→SE4	-.16 [-.41; .09]	-.03 [-.31; .25]	**-.38** **[-.68; -.08]**	**-.45** **[-.81; -.09]**	.00 [-.25; .24]	.09 [-.22; .39]	-.05 [-.30; .20]	.16 [-.77; 1.09]

Note. Exploratory study: N_{Total} = 683; N_A = 179; N_B = 163; N_C = 170; N_D = 171. Confirmatory study: N_{Total} = 1,087; N_A = 280; N_B = 269; N_C = 273; N_D = 265. Condition A: Self-esteem trait, Depressive symptoms state; Condition B: Self-esteem trait, Depressive symptoms trait; Condition C: Self-esteem state, Depressive symptoms trait; Condition D: Self-esteem state, Depressive symptoms state. Bold indicates that the 95% confidence interval did not include zero. Italics indicate that the 90% confidence interval did not include zero. All coefficients are standardized.

5 STUDY 3: IS TEACHER ATTACHMENT PROSPECTIVELY RELATED TO SELF-ESTEEM? A 10-YEAR LONGITUDINAL STUDY OF MEXICAN-ORIGIN YOUTH

Braun, L., Lawson, K., Spengler, M., Göllner, R., & Robins, R. W. (2020). Is teacher attachment prospectively related to self-esteem? A 10-year longitudinal study of Mexican-origin Youth. *Manuscript submitted for publication.*

Abstract

Adolescents spend much of their day-to-day life at school, where they continuously interact with teachers, who often provide them with both academic and emotional support. Although previous research and theory suggests that close emotional connections can foster self-esteem, the role of teacher attachment in students' self-esteem development has yet to be examined. The present study used data from 674 Mexican-origin youth followed annually from age 11 to 21 to test eight preregistered hypotheses about reciprocal relations between teacher attachment (assessed via student-perceived teacher support and teacher rejection) and students' global self-esteem (assessed via the Rosenberg Self-Esteem scale). Results showed (1) positive prospective reciprocal relations between teacher support and self-esteem, based on both cross-lagged panel models (CLPM) and random intercept cross-lagged panel models (RI-CLPM), and (2) negative prospective reciprocal relations between teacher rejection and self-esteem, based only on CLPMs. Contrary to expectations, neither school importance nor family support moderated the prospective relations between teacher support/rejection and self-esteem. Additionally, a trait factor of teacher support/rejection from age 11 to 16 predicted self-esteem at age 21, even after controlling for self-esteem at age 11 (but not controlling for self-esteem at age 16). Discussion focuses on the importance of teacher attachment in students' self-esteem development and the transactional processes that underlie reciprocal relations between teacher attachment and self-esteem.

Keywords: self-esteem, teacher attachment, adolescence, (random intercept) cross-lagged panel models

Is Teacher Attachment Prospectively Related to Students' Self-Esteem? A 10-Year Longitudinal Study of Mexican-origin Youth

Children and adolescents have almost daily contact with their teachers, who can serve as important attachment figures by providing them with a sense of emotional security. Previous research has found a strong link between positive interpersonal relationships and self-esteem, but questions remain about the direction of this effect and its generalizability to other contexts such as school (Harris & Orth, 2019). On the one hand, individuals with higher self-esteem tend to have stronger feelings of attachment towards others, leading to more fulfilling relationships (e.g., Bowlby, 1969; Leary & Baumeister, 2000). On the other hand, individuals experiencing more fulfilling relationships might go on to develop higher self-esteem. However, almost no work has examined directional influences between teacher attachment and self-esteem. The present study aims to fill this gap by analyzing reciprocal relations between teacher attachment (assessed via student-perceived teacher support and teacher rejection) and students' global self-esteem (assessed via the Rosenberg Self-Esteem scale) using data from a community sample of 674 Mexican-origin youth followed longitudinally from age 11 to 21.

Self-esteem Development in Adolescence

Self-esteem refers to an individual's "subjective evaluation of his or her worth as a person" (Donnellan, Trzesniewski, & Robins, 2011, p. 718). Past research has found that self-esteem shows relatively high rank-order stability across the lifespan (Trzesniewski, Donnellan, & Robins, 2003), but also well-documented mean-level changes across different developmental periods (Orth & Robins, 2019). In line with the maturity principle of personality development (Roberts & Wood, 2006), adolescence is a period characterized by average *increases* in self-esteem (Orth, Erol, & Luciano, 2018). However, some studies have documented temporary *declines*, especially after the transition to junior high-school (e.g., Eccles & Midgley, 1989; Wigfield, Eccles, Mac Iver, Reuman, & Midgley, 1991; Wigfield & Eccles, 1994; but see Wagner, Lüdtke, Robitzsch, Göllner, & Trautwein, 2018), which may be a consequence of disruption during puberty (Soto & Tackett, 2015), increasing self-awareness (Rosenberg, Schooler, & Schoenbach, 1989), or increasing exposure to negative feedback from teachers, parents, and peers (Robins, Tracy, Trzesniewski, Potter, & Gosling, 2001). Thus, although there are well-documented changes in self-esteem during adolescence, we know relatively little about the factors that influence self-esteem change during this period, especially with regard to factors in the school context.

Teacher Attachment

Schools are vitally important for students' development (Eccles & Midgley, 1989). In particular, teachers have substantial power to shape students' experiences at school (Hamre & Pianta, 2001). Over and above teaching course content and managing classroom situations, many teachers seek to create positive and constructive relationships with their students by encouraging them academically and providing an emotionally supportive climate (Wentzel, 2009). Conversely, many students strive to receive emotional support and acceptance from their teachers (Kesner, 2000). This academic and emotional support means that teachers can be the most important non-familial person in students' lives (Kesner, 2000), highlighting the importance of teacher attachment for positive youth development.

We conceptualize teacher attachment as a global term that comprises two dimensions –teacher support and teacher rejection. *Teacher support* involves students' perceptions of being understood and supported by their teachers concerning academic and personal problems. *Teacher rejection* encompasses students' perception of being criticized, disliked, and disrespected by their teachers. Despite the conceptual relation between these two dimensions, we do not consider them to be opposite ends of the same continuum. Rather, students may perceive high (or low) levels of teacher support and high (or low) levels of teacher rejection simultaneously. For example, one can imagine a situation in which a student does not feel actively supported by his/her teachers, but at the same time does not feel actively rejected. Distinguishing between teacher support and teacher rejection is consistent with previous research that differentiates positive (e.g., closeness) and negative aspects (e.g., conflict) of teacher attachment (Hamre & Pianta, 2001; Pianta & Steinberg, 1992). Furthermore, past factor analytic work has shown that these two dimensions form independent factors (Hamre & Pianta, 2001), which fits with the general idea of differentiating positive and negative affective attributes (Watson, Clark, & Tellegen, 1988). Relatedly, negative experiences with teachers seem to be particularly related to problem behavior, whereas positive experiences are more closely related to school engagement and school adjustment (Hamre & Pianta, 2001; Martin & Collie, 2019).

Students' experiences in the classroom are often described as outcomes of multiple offer-use situations (Creemers, Kyriakides, & Antoniou, 2013; Helmke, 2012; Scheerens & Bosker, 1997). From this viewpoint, teacher attachment can be framed as an offer to students, which in turn has to be perceived and then used by the students themselves in order to impact

students' feelings and behavior. Thus, measures of teacher attachment based on reports by others (e.g., teachers) may not be the most relevant indicators of students' social-emotional development. Instead, students' own perceptions of teacher attachment are likely to influence their social-emotional development (Aldrup, Klusmann, Lüdtke, Göllner, & Trautwein, 2018; Furrer & Skinner, 2003; Kunter & Baumert, 2006). This perception might be particularly important for socio-emotional constructs like self-esteem, given its dependency on subjective, affective evaluations of personal experiences. Additionally, especially during adolescence, students usually have multiple teachers at a time, meaning it can be useful to focus on an overall, cross-situational assessment of perceived teacher attachment (versus focusing on a single relationship between the student and a particular teacher) when the research question is aimed at understanding effects on global outcomes (e.g., self-esteem).

Like self-esteem, students' attachment to their teachers might undergo changes, especially during the transitional phase of adolescence. Adolescence is a challenging developmental period, marked by physical and emotional changes, an increasing need for independence, and a search for identity and subsequent reformulation of perceptions and evaluations of self (Arnett, 1999; Block & Robins, 1993; Erikson, 1968; Hall, 1904). Along with shifts in motivational constructs during this time (e.g., Eccles et al., 1993; Scherrer & Preckel, 2019), some studies suggest that students' perceptions of their teachers become more negative across childhood and adolescence (Castro-Schilo, Ferrer, Hernández, & Conger, 2016; Eccles & Roeser, 2011; Hughes & Cao, 2018; Lynch & Cicchetti, 1997; Reddy, Rhodes, & Mulhall, 2003).

Association between Teacher Attachment and Self-Esteem

Multiple theoretical perspectives suggest that self-esteem is closely tied to interpersonal relations (Bowlby, 1969; Cooley, 1902; Deci & Ryan, 1985; Leary & Baumeister, 2000; Orth & Robins, 2019). In fact, self-esteem was originally defined as a self-evaluation that is reflected through the eyes of significant (i.e., close) others (Cooley, 1902). However, self-esteem is not purely a measure of how individuals are perceived by others, but rather a reflection how individuals *evaluate* their relationships with others. This stems from the basic human "need to belong" (Deci & Ryan, 1985), which, when unfulfilled, can result in lower levels of self-esteem (Eccles & Midgley, 1989). In particular, sociometer theory hypothesizes that self-esteem represents one's perceived relational value, which is based on people's perception of how positively they are perceived by others (Leary, Tambor, Terdal, & Downs, 1995; Leary & Baumeister, 2000). From this perspective, self-esteem changes when individuals perceive shifts in their

relational value. Relational perceptions can arise from feelings of secure attachment, which early in life shape the internal representations of feeling accepted and valued by others (Bowlby, 1969, 1973, 1980). Previous research often distinguished between positive and negative relational experiences as drivers of ones' self-esteem. Thus, drawing on meta-analytical evidence, positive interpersonal experiences (i.e., acceptance) were more strongly related to self-esteem than negative interpersonal experiences (i.e., rejection; Blackhart, Nelson, Knowles, & Baumeister, 2009). However, it is unlikely that every positive or negative social experience is crucial for one's self-esteem. Rather, self-esteem is particularly affected by social relations with *significant others* (Cooley, 1902; James, 1890/1963; Leary & Baumeister, 2000). As such, parents who represent the primary significant attachment figures set the individuals' internal working model for feelings of security, care, and love (Bowlby, 1969; Kesner, 2000).

Previous empirical research supports the theory that interpersonal relationships and self-esteem are related over time (Harris & Orth, 2019). For example, parent-child relationship quality is positively related to self-esteem both cross-sectionally (e.g., Barber, Chadwick, & Oerter, 1992; Verschueren, Marcoen, & Schoefs, 1996; Whitbeck et al., 1991) and longitudinally (e.g., Boudreault-Bouchard et al., 2013; Krauss, Orth, & Robins, 2019; Orth, 2018; but see Harris et al., 2015). In addition to research on parents, past work has highlighted the importance of peers in shaping an individual's self-esteem (Gruenenfelder-Steiger, Harris, & Fend, 2016; Reitz, Motti-Stefanidi, & Asendorpf, 2016; Wagner et al., 2018). However, the role of teachers has been largely ignored, despite teachers being important attachment figures in students' lives who might provide a compensatory resource when students lack other attachment figures (Kesner, 2000).

Multiple longitudinal studies suggest that different aspects of teacher-student relationships predict students' educational and psychological outcomes, such as self-perceived competence in specific school subjects (Rice et al., 2013; Scherer, Nilsen, & Jansen, 2016; Wagner et al., 2016), increased engagement and motivation (Midgley, Feldlaufer, & Eccles, 1989; Roorda, Koomen, Spilt, & Oort, 2011; Ruzek et al., 2016; Scherer et al., 2016), and improved psychological adjustment (Kunter & Voss, 2013; Pössel, Rudasill, Sawyer, Spence, & Bjerg, 2013). Indeed, one longitudinal study found that trajectories of teacher support and self-esteem were positively related during middle school (Reddy et al., 2003).

Given the strong theoretical foundation about the impact of interpersonal relationships on self-esteem (e.g., Leary & Baumeister, 2000) and the socio-emotional role of teachers in the

school context (Kesner, 2000), we predict that teacher attachment will have a positive prospective effect on students' self-esteem. In particular, students who believe they have a positive relationship with their teachers might experience increases in their relational value, feel more secure about their worth as a student, and have a stronger sense of belonging, all of which might contribute to increases in self-esteem. Further, the relation between teacher attachment and self-esteem is likely a transactional process, characterized by reciprocal associations between the two constructs (Sameroff, 2009; Swann, Chang-Schneider, & Larsen McClarty, 2007). For example, high self-esteem students might project beliefs about the self to beliefs about relationships to teachers (Murray, Holmes, & Griffin, 2000) or these high self-esteem students might broadcast themselves more in the classroom and therefore garner more attention, be more involved, and be more well-liked (Srivastava & Beer, 2005), which can result in more positive attachment to teachers. At the same time, students with low self-esteem might respond more intensively to negative interpersonal experiences, such as critical feedback from teachers (Nezlek, Kowalski, Leary, Blevins, & Holgate, 1997), which might impede the development of a positive attachment to teachers.

We know of no longitudinal studies that have examined whether teacher attachment prospectively predicts subsequent levels of self-esteem, and vice versa. Consistent with this view, a recent meta-analysis of reciprocal associations between social relationships and self-esteem did not include any studies involving student-teacher relationships, even though "teacher" was included as a search term (Harris & Orth, 2019). This gap in the literature highlights the need for research examining longitudinal reciprocal relations between teacher attachment and self-esteem. Furthermore, given self-esteem's numerous longterm beneficial consequences (Orth, Robins, & Widaman, 2012), it is particularly important to investigate whether factors such as teacher attachment promote healthy self-esteem development during the critical adolescent years and into young adulthood. Thus, in addition to examining prospective associations between teacher attachment and self-esteem during the school years, we want to investigate whether the overall experience of teacher attachment during the school years had long-term effects on self-esteem several years after high school.

The strength of the association between teacher attachment and self-esteem may be impacted by a number of factors. In the present study, we focus on two such factors: importance of school and family emotional support. Many self-esteem theories emphasize that experiences in domains that are more highly valued by an individual will have a greater impact on that individual's self-esteem (James, 1890/1963). From this perspective, teacher attachment should

have an *amplified* influence on the self-esteem of students who value the school context, compared to those who do not value the school context. Moreover, when youth lack other relational sources of self-esteem, such as positive attachment to their families, teacher attachment might serve as a *compensatory* resource that plays a particularly significant role in bolstering self-esteem (Burchinal, Peisner-Feinberg, Pianta, & Howes, 2002; Mitchell-Copeland, Denham, & DeMulder, 1997). From this perspective, teacher attachment should have a stronger influence on the self-esteem of students who feel less supported by their families, compared to those who receive more family support. Further supporting the possibility that lack of family support might magnify the association between teacher attachment and self-esteem, youth with an adverse history of family attachment might experience more negative feelings about the self when negative attachment experiences are confirmed by other important attachment figures, such as their teachers (McGrath & van Bergen, 2015; Ryan, Stiller, & Lynch, 1994). Therefore, it is important to consider students' beliefs about the importance of school as well as the support from their family when addressing research questions about teacher attachment and self-esteem.

The Present Study

The present study investigates the relation between perceived teacher attachment and self-esteem in a longitudinal study of Mexican-origin youth assessed repeatedly from age 11 to 21. In particular, we examined whether, and to what extent, teacher attachment and self-esteem are prospectively and reciprocally related over time. We consider both positive (teacher support) and negative (teacher rejection) aspects of teacher attachment, consistent with past research (e.g., Martin & Collie, 2019; Pianta & Steinberg, 1992). Additionally, we evaluate our research questions using data from a non-WEIRD sample (i.e., Mexican-origin youth living in California), which is important because ethnic minority youth are at particular high risk for adverse social, emotional, and academic outcomes (Brey et al., 2019; OECD, 2016, 2018). Moreover, feelings of attachment to teachers might provide an emotional anchor for these students by supporting them in the face of socioeconomic, sociolinguistic, and acculturation-related challenges (Eccles & Roeser, 2011; Garcia-Reid, Reid, & Peterson, 2005; Roorda et al., 2011). Beyond its theoretical and practical relevance, there is an emerging need in psychology to analyze data from non-WEIRD or disadvantaged samples, since these samples have been traditionally underrepresented in psychological research (Furrer & Skinner, 2003; Henrich, Heine, & Norenzayan, 2010).

The present study examined three broad research questions. The hypotheses and the research plan were preregistered at the Open Science Framework at the following link: https://osf.io/r86b5/?view_only=fe0f9c6494b64250940c6d7d2a6f57c1. We used data for all available participants (i.e., no exclusions were applied), and we have reported all analyses conducted to address our research questions.

First, given the importance of social experiences in the development of self-esteem (e.g., Leary & Baumeister, 2000), as well as the particular role of teachers as attachment figures in students daily experiences (e.g., Kesner, 2000), we tested whether student-perceived teacher attachment was prospectively related to students' self-esteem from age 11 to 16 (Research Question 1). In addition, we explored whether beliefs in the importance of school and student-reported support from their family serve as amplifying or compensatory drivers of this relation. We formulated two hypotheses about the prospective effect of teacher attachment (teacher support and rejection) on self-esteem (H1.1 and H1.2) and two hypotheses about possible moderators of this relation (H1.3 and H1.4):

- H1.1: We predicted a *positive* prospective association between students' perceptions of teacher support and students' self-esteem, controlling for their prior levels of self-esteem.
- H1.2: We predicted a *negative* prospective association between students' reports of teacher rejection and students' self-esteem, controlling for their prior levels of self-esteem.
- H1.3: We predicted that the associations specified in Hypotheses 1.1. and 1.2 will be moderated by students' reports of how important school is to them. Specifically, we expected that the hypothesized associations between teacher attachment and self-esteem would be *stronger* (i.e., more positive for Hyp. 1.1 and more negative for Hyp. 1.2) when students believe that school is more (vs. less) important (Amplification Hypothesis).
- H1.4: We predicted that the associations specified in Hypotheses 1.1. and 1.2 will be moderated by students' perceived support from their families. Specifically, we expected that the hypothesized associations between teacher attachment and self-esteem would be *stronger* (i.e., more positive for Hyp. 1.1 and more negative for Hyp. 1.2) when students feel less (vs. more) supported by their families (Compensation Hypothesis).

Second, we examined whether self-esteem was prospectively related to teacher attachment from age 11 to 16 (Research Question 2). In other words, we examined whether the relation between self-esteem and teacher attachment follows a transactional process, in which not

only is teacher attachment related to students' self-esteem, but also students' self-esteem affects teachers' attachment behaviors and eventually students' perception of this attachment.

- H2.1: We predicted that self-esteem will have a *positive* prospective association with teacher support, controlling for prior teacher support.
- H2.2: We predicted that self-esteem will have a *negative* prospective association with teacher rejection, controlling for prior teacher rejection.

Third, to examine whether teacher attachment has enduring consequences for students' self-esteem beyond the school years, we investigated whether teacher attachment assessed from age 11 to age 16 is prospectively associated with self-esteem assessed four years after high school[16] (i.e., age 21; Research Question 3):

- H3.1: We predicted that the trait factor of teacher support from age 11 to age 16 would have a *positive* association with self-esteem four years post-high school, such that, on average, students who report higher levels of teacher support from age 11 to 16 will have higher self-esteem at age 21. We tested three (pre-registered) models: 1) not controlling for prior self-esteem, 2) controlling for level of self-esteem at age 11[17] and 3) controlling for level of self-esteem at age 16.
- H3.2: We predicted that the trait factor of teacher rejection from age 11 to 16 would have a *negative* association with self-esteem four years post-high school, such that, on average, students who report higher levels of teacher rejection from age 11 to 16 will have lower self-esteem at age 21. We tested three different (pre-registered) models: (1) not controlling for prior self-esteem, (2) controlling for level of self-esteem at age 11, and (3) controlling for level of self-esteem at age 16.

In light of the current debate about different models of cross-lagged effects (e.g., Orth, Clark, Donnellan, & Robins, in press; Usami, Murayama, & Hamaker, 2019; Zyphur et al., 2019), we used two different statistical models to test the hypotheses based on Research Question 1 and 2. First, we applied traditional cross-lagged panel models (CLPM), which allow us

[16] In the preregistration, we mistakenly described the final wave of self-esteem data collection as *three* years post-high school with a median age of 20 years, but this wave is actually four years post-high school with a median age of 21 years (because the time interval between waves shifts from annual data collection across the first ten waves to biennial data collection from Wave 10 to 11).

[17] In the 6th grade assessment (age 11), data were missing for approximately half the sample because the RSE was added to the study halfway through the wave of data collection. We repeated the analyses addressing Research Question 3 controlling for self-esteem at age 12. All significant result remained significant and the magnitude of the effects were very similar to those found when controlling for self-esteem at age 11.

to investigate effects based on between-person differences. A significant cross-lagged effect in this model indicates that students with a more positive attachment to their teachers tend to show subsequent rank-order increases in their self-esteem compared to students with a less positive attachment to their teachers. Second, we used random intercept cross-lagged panel models (RI-CLPM; Hamaker, Kuiper, & Grasman, 2015) to investigate effects based on within-person deviations. In contrast to the CLPM, the RI-CLPM explicitly models the stable between-person variance (i.e., trait variance) for each construct. Consequently, a cross-lagged effect in this model tests for the prospective effect of a within-person deviation from the trait level of one construct (e.g., teacher attachment) on change in the within-person deviation from the trait level of the other construct (e.g., student self-esteem). For example, a positive cross-lagged effect from teacher attachment to student self-esteem indicates that when a student has a more positive teacher attachment than usual at a particular time point, the student will have higher self-esteem than usual at a subsequent time point. Note that existing theories of self-esteem do not provide the level of precision required to make differential predictions about the existence of between vs. within-person effects of teacher attachment on self-esteem (or self-esteem on teacher attachment). Consequently, all hypotheses are predicted to hold at both the between and within-person level.

Methods

Participants and Procedures

Data come from the California Families Project, an ongoing longitudinal study of Mexican-origin youth and their parents (N = 674)[18]. Children were drawn at random from rosters of students from the Sacramento and Woodland, CA, school districts. The focal child had to be in the 5th grade, of Mexican origin, and living with his or her biological mother, in order to participate in the study. Approximately 72.6% of the eligible families agreed to participate in the study, which was granted approval by the University of California, Davis Institutional Review Board (Protocol # 217484-21). The children (50% female) were assessed annually from 5th grade (M_{age}=10.86, SD_{age}=0.51) to two years post-high school and then again two years later, or four years post-high school (M_{age}=21.74, SD_{age}=0.73). Data collection occurred from 2007 to 2019. For the present study, we used data from the 6th to 11th grade assessments and the 4 years post-high school assessment, when the key study variables were available.

[18] Please note that intentionally the description of the sample is partly identical with previous publications on the California Families Project. For a full list of California Families Project publications, see: https://osf.io/ky7cw/

Participants were interviewed in their homes in Spanish or English, depending on their preference. Interviewers were all bilingual and most were of Mexican heritage. Sixty-three percent of mothers and 65% of fathers had less than a high school education (median = 9th grade education level for both mothers and fathers). Median total household income was $32,500 at Wave 1. With regard to generational status, 83.6% of mothers and 89.4% of fathers were 1st generation immigrants. Retention rates (relative to the original sample of 674 youth) were: 85% (age 11), 86% (Age 12), 88% (Age 13), 90% (Age 14), 88% (Age 15), 89% (Age 16), and 80% (Age 21). We tested attrition effects on the main study variables (teacher support, teacher rejection, self-esteem, school importance, family support) and found no differences between continuers and those who dropped out from the study early (see Table S1 in the Supplemental Material).

Measures

Self-esteem. Children reported their global self-esteem annually from age 11 to age 21 using the 10-item Rosenberg Self-Esteem Scale (RSE; Rosenberg, 1965). Items from the RSE include "On the whole, I am satisfied with myself" and "I feel that I have a number of good qualities," which were rated on a scale ranging from 1 (*Strongly disagree*) to 4 (*Strongly agree*). For Research Questions 1 and 2, we used annual RSE data from age 11 to age 16 (6 waves of data). For Research Question 3, we additionally included RSE data assessed four years post-high school (age 21). Reliability (ω) of the RSE ranged from .78 to .90 (see Table S2 in the Supplemental Material).

Teacher support. Children completed a measure of *teacher support* annually from age 11 to age 16 using nine items adapted from the Inventory of Parent and Peer Attachment (Armsden & Greenberg, 1987). For each item, children reported the frequency with which an event had occurred during the past three months. These items include, "A teacher helped you with your problems" and "You could count on a teacher when you needed to talk." All items were rated on a scale ranging from 1 *(Almost never or never*) to 4 *(Almost always or always*). Reliability (ω) of the measure ranged from .85 to .93 (see Table S2 in the Supplemental Material).

Teacher rejection. Children completed a measure of *teacher rejection* annually from age 11 to age 16 using a four-item scale that was developed for the California Families Project. These items include, "Your teachers do not like you very much" and "Your teachers usually put you down," which were rated on a scale ranging from 1 (*Not at all true*) to 4 (*Very true*). Reliability (ω) ranged from .58 to .71 (see Table S2 in the Supplemental Material).

Importance of school. Children reported on their beliefs about the *importance of school* annually from age 11 to age 16 using a seven-item scale that was designed to assess the utility value, attainment value, and intrinsic value of school (Roeser, Lord, & Eccles, 1994; Smith et al., 1997). These items include, "You like to do well in school," "It is important to finish high school," and "You like school a lot," which were rated on a scale ranging from 1 (*Not at all*) to 4 (*Very true*). Reliability (ω) ranged from .71 to .80 (see Table S2 in the Supplemental Material).

Family support. Children reported their level of perceived *family support* age 12 and 14 using four items from the Multidimensional Support Scale (Zimet, Dahlem, Zimet, & Farley, 1988). These items include "Your family really tries to help you" and "You can talk about your problems with your family," which were rated on a scale ranging from 1 (*Not at all true*) to 4 (*Very true*). Reliability (ω) ranged from .88 to .91 (see Table S2 in the Supplemental Material).

Statistical Analyses

All statistical analyses were carried out in the framework of longitudinal structural equation modeling using Mplus Version 8 (Muthén & Muthén, 1998-2017). All data analysis scripts are available for review at the Open Science Framework at the following address: https://osf.io/7hkdp/?view_only=543a38c96565470d8e132fae58a96c5b (the project will be publicly available after the review process has been completed). The alpha level was set to .05 (two-tailed) and we reported 95% confidence intervals. In all models, we accounted for missing data by using full-information maximum likelihood (FIML) estimation (Allison, 2003; Schafer & Graham, 2002). For self-esteem, teacher support, and school importance, we built three item parcels each because item parcels can lead to more reliable latent variables than individual items (Little, Cunningham, Shahar, & Widaman, 2002; but see Marsh, Lüdtke, Nagengast, Morin, & Davier, 2013).We did not use item parcels for the teacher rejection and family support scales because there were too few items. Indicators based on the same items were correlated across waves (Cole, Ciesla, & Steiger, 2007). Absolute model fit was assessed with the comparative fit index (CFI), the Tucker–Lewis index (TLI), and the root-mean-square error of approximation (RMSEA). Specifically, good fit was indicated by CFI and TLI values greater than or equal to .95 and RMSEA values less than or equal to .06 (Hu & Bentler, 1999). We assessed differences in model fit via change in comparative fit index (ΔCFI) less than or equal to .01, change in Tucker-Lewis index (ΔTLI) less than or equal to .01 (Chen, 2007), change in

McDonald's non-centrality index (ΔNCI) less than or equal to .02, and change in chi-square and degrees of freedom (Cheung & Rensvold, 2002; Meade, Johnson, & Braddy, 2008).

Longitudinal measurement invariance. Before testing our hypotheses, we analyzed longitudinal measurement invariance of latent self-esteem and both latent teacher attachment measures (i.e., teacher support and teacher rejection) to examine whether these constructs were measured similarly over time. To evaluate this, we compared three measurement models (Meredith, 1993): (1) freely estimating the factor loadings for the latent factors over time (i.e., configural invariance); (2) constraining the respective factor loadings to be equal over time (i.e., weak invariance); and (3) constraining the factor loadings and intercepts to be equal over time (i.e., strong invariance).

Cross-lagged regression models. To test Hypotheses 1.1, 1.2, 1.3, 1.4, 2.1, and 2.2, we conducted latent variable CLPMs to examine the between-person reciprocal relations between teacher attachment and self-esteem, and teacher rejection and self-esteem. In a CLPM, the cross-lagged paths indicate the prospective relations between self-esteem on teacher support/rejection (and vice versa) from lag to lag, after controlling for both constructs' stability over time. For example, a positive cross-lagged coefficient from teacher support to self-esteem in the CLPM would indicate that students with higher teacher support than other students are predicted to have higher self-esteem than other students at the next time point (having controlled for previous self-esteem). We tested the fit of three structural models: (1) a model that allows all paths to be freely estimated, (2) a model where the autoregressive paths were constrained to be equal over time within each construct, and (3) a model where the autoregressive paths and the cross-lagged paths were constrained to be equal over time within each construct. If the differences in fit indices between these specifications were not significant, then we favored the more parsimonious model and retained the structural constraints.

Random intercept cross-lagged regression models. In addition to testing Hypotheses 1.1, 1.2, 1.3, 1.4, 2.1, and 2.2 using a CLPM, we also analyzed the data using random intercept cross-lagged regression models (RI-CLPM; Hamaker et al., 2015). The RI-CLPM allows us to examine between-person reciprocal relations on within-person deviations between teacher support/rejection and self-esteem. In comparison the CLPM, the RI-CLPM controls for stable between-person differences and thus, estimates autoregressive and cross-lagged coefficients based on the within-person deviations from the individuals' typical score. For example, a pos-

itive cross-lagged coefficient from teacher support to self-esteem in the RI-CLPM would indicate that students whose deviations from their typical teacher support are higher tend to have higher deviations from their typical self-esteem at the next time point (controlling for the previous within-person deviation in self-esteem). As with the CLPMs, we tested the fit of three RI-CLPMs: (1) a model that allows all paths to be freely estimated, (2) a model where the autoregressive paths were constrained to be equal over time within each construct, and (3) a model where the autoregressive paths and the cross-lagged paths were constrained to be equal over time within each construct. If the difference in fit indices between these specifications was not significant, then we favored the more parsimonious model and retained the structural constraints.

Interaction effects. To test Hypotheses 1.3 and 1.4, we examined whether school importance and family support moderate the effect of teacher support/rejection on self-esteem. We therefore included the main effects of school importance/family support as well as the interaction effects between school importance/family support and teacher support/rejection in the CLPMs. The interaction effects were modeled as latent interactions using the latent moderated structural (LMS) equations approach (Klein & Moosbrugger, 2000). Due to the complexity of the models, we run latent interactions separately for all available waves (school importance at age 11 to 15; family support at age 12 and 14). We standardized the latent variables before modeling the latent interaction terms in order to decrease the risk of multicollinearity. In addition, in order to make the FIML estimation comparable across the models, we included all autoregressive and cross-lagged paths at all time-points in all models. Finally, for these models, we did not constrain autoregressive and cross-lagged paths to be equal across time-points because they represented different partial regression coefficients when the moderation effect was included compared to when the moderation effect was not included.

Latent state-trait model. For testing Hypotheses 3.1 and 3.2**,** we modeled a latent state-trait model (Steyer, Schmitt, & Eid, 1999) separately for teacher support and teacher rejection from age 11 to 16. Subsequently, we predicted self-esteem four years post-high school by the trait factor of teacher support/rejection. We tested three (pre-registered) models: (1) not controlling for prior self-esteem (2) controlling for self-esteem at age 11, and (3) controlling for self-esteem at age 16.

Multiple-group analyses. To test whether the effects were robust across gender, we conducted exploratory multiple-group analyses separately for all models.

Results

For all main study variables (i.e., self-esteem, teacher support, teacher rejection, school importance, family support) we observed strong measurement invariance over time. In other words, model fit did not decrease substantially when we constrained loadings and intercepts to be equal compared to models with weak or configural invariance (for details see Table S3 in the Supplemental Material). Descriptives and rank-order stabilities of the models are presented in Table S2 in the Supplemental Material. In the (RI-) CLPMs, constraining the autoregressive and lagged paths to be equal across time points did not decrease the model fit substantially relative to freely estimated models (for details see Table S4 in the Supplemental Material). Therefore, we report standardized coefficients from models in which the autoregressive and cross-lagged effects were constrained to be equal across time. For transparency, we provide the results of the unconstrained models in Tables S5 and S6 in the Supplemental Material.

Table 1 presents the results of the standardized estimates of the (random intercept) cross-lagged panel models. All autoregressive coefficients from CLPMs for self-esteem (β = .61 – 67), teacher support (β = .47 – .57), and teacher rejection (β = .51 – 65) were statistically significant (see Table 1). In the RI-CLPM, the autoregressive coefficients on the within person deviations of self-esteem (β = .34 – 40) and teacher support (β = .25 – 34) were statistically significant, however they were not for teacher rejection (β = .19 – 27; see Table 1).

Is Teacher Attachment Prospectively Associated with Self-Esteem?

Table 1 depicts the standardized cross-lagged effects of teacher support/rejection on self-esteem based on the CLPM (i.e., between-person effects) and the RI-CLPM (i.e., within-person deviations). Teacher support positively predicted changes in self-esteem in both the CLPM (β = .06 – 08) and RI-CLPM (β = .05 –.07). Moreover, teacher rejection negatively predicted changes in self-esteem in the CLPM (β = -.04 – -.05), but not in the RI-CLPM.

Table 1

Standardized Parameter Estimates [and 95% Confidence Intervals] of the (Random Intercept) Cross-Lagged Panel Model (CLPM; RI-CLPM) between Self-Esteem and Teacher Attachment (Support and Rejection)

	Self-Esteem and Teacher Support		Self-Esteem and Teacher Rejection	
Path	CLPM	RI-CLPM	CLPM	RI-CLPM
Autoregres-sive				
TA11→TA12	**.47 [.42; .52]**	**.25 [.19; .31]**	**.51 [.33; .68]**	.19 [-.27; .64]
TA12→TA13	**.53 [.48; .58]**	**.30 [.22; .37]**	**.59 [.45; .73]**	.26 [-.32; .83]
TA13→TA14	**.54 [.49; .60]**	**.30 [.21; .38]**	**.65 [.49; .81]**	.27 [-.35; .89]
TA14→TA15	**.54 [.49; .60]**	**.31 [.23; .39]**	**.64 [.45; .83]**	.23 [-.38; .83]
TA15→TA16	**.57 [.52; .62]**	**.34 [.25; .42]**	**.60 [.47; .72]**	.21 [-.35; .78]
SE11→SE12	**.62 [.57; .68]**	**.38 [.27; .49]**	**.62 [.56; .68]**	**.40 [.28; .51]**
SE12→SE13	**.66 [.60; .73]**	**.36 [.22; .50]**	**.67 [.61; .73]**	**.36 [.22; .51]**
SE13→SE14	**.61 [.56; .67]**	**.34 [.23; .45]**	**.62 [.57; .68]**	**.35 [.24; .46]**
SE14→SE15	**.66 [.60; .72]**	**.37 [.22; .51]**	**.67 [.61; .73]**	**.38 [.23; .54]**
SE15→SE16	**.63 [.58; .68]**	**.36 [.23; .48]**	**.64 [.59; .69]**	**.36 [.23; .50]**
Cross-lagged				
SE11→TA12	**.13 [.08; .17]**	**.10 [.02; .18]**	**-.06 [-.11; -.02]**	.01 [-.17; .18]
SE12→TA13	**.12 [.08; .17]**	**.09 [.01; .17]**	**-.07 [-.12; -.02]**	.01 [-.17; .19]
SE13→TA14	**.12 [.08; .16]**	**.08 [.01; .15]**	**-.07 [-.12; -.02]**	.01 [-.21; .23]
SE14→TA15	**.12 [.08; .16]**	**.09 [.01; .16]**	**-.08 [-.12; -.03]**	.01 [-.22; .25]
SE15→TA16	**.12 [.08; .16]**	**.09 [.01; .17]**	**-.07 [-.12; -.03]**	.01 [-.20; .23]
TA11→SE12	**.06 [.03; .09]**	**.05 [.00; .10]**	**-.04 [-.07; -.01]**	.06 [-.07; .18]
TA12→SE13	**.07 [.04; .10]**	**.07 [.00; .13]**	**-.05 [-.09; -.01]**	.06 [-.07; .20]
TA13→SE14	**.07 [.04; .11]**	**.07 [.00; .13]**	**-.05 [-.09; -.01]**	.06 [-.07; .18]
TA14→SE15	**.08 [.04; .11]**	**.07 [.00; .14]**	**-.05 [-.09; -.01]**	.05 [-.06; .16]
TA15→SE16	**.08 [.04; .12]**	**.07 [.00; .15]**	**-.04 [-.08; .00]**	.05 [-.06; .15]

Note. *N*=674; SE = Self-esteem; TA= Teacher attachment (support or rejection). Bold indicates that the 95% confidence interval did not include zero.

These findings are in line with our hypothesis (H1.1) that teacher support would predict later self-esteem, and partially in line with our hypothesis (H1.2) that teacher rejection would predict later self-esteem.

Amplification hypothesis. To test whether beliefs in the importance of school amplified the effect of teacher attachment on self-esteem, we included latent moderated terms between teacher support/rejection and school importance. As shown in Table 2, neither the interaction between teacher support and school importance, nor the interaction between teacher rejection and importance was statistically significant. These findings are not consistent with the amplification hypothesis (H1.3).

Compensatory hypothesis. Next, we tested whether the effect of teacher attachment on self-esteem could compensate for students' perceived lack of family support using latent moderated terms between teacher support/rejection and family support. Table 2 shows that neither the interaction between teacher support and family support, nor the interaction between teacher rejection and family support was statistically significant. Hence, there was no support for the compensatory hypothesis (H1.4).

Is Self-Esteem Prospectively Associated with Teacher Attachment?

In addition to the prospective effect from teacher attachment to self-esteem, we were also interested in the effect from self-esteem to teacher attachment. Table 1 shows the standardized cross-lagged coefficients and 95 % confidence intervals based on the CLPM and the RI-CLPM. Self-esteem positively predicted changes in teacher support in both the CLPM (β = .12 – .13) and RI-CLPM (β = .08 – .10); Moreover, self-esteem negatively predicted changes in teacher rejection in the CLPM (β = -.06 – -.08), but not in the RI-CLPM. These findings are in line with our hypothesis (H2.1) that self-esteem would predict teacher support, and partially in line with our hypothesis (H2.2) that self-esteem would predict later teacher rejection.

Is Teacher Attachment from Age 11 to 16 Associated with Self-Esteem Four Years after High-School?

Next, we investigated whether the prospective association between teacher attachment and self-esteem held up when self-esteem was assessed four years post-high school. As shown in Table 3, a trait factor of teacher support, which represented the common variance of teacher

Table 2

Standardized Parameter Estimates [and 95% Confidence Intervals] for Predicting Self-Esteem by Latent Moderated Terms within Cross-Lagged Panel Models

Predictor	Teacher support					Teacher rejection				
	11 to 12 years	12 to 13 years	13 to 14 years	14 to 15 years	15 to 16 years	11 to 12 years	12 to 13 years	13 to 14 years	14 to 15 years	15 to 16 years
School Importance										
TA → SE	-.05 [-.18; .08]	.03 [-.08; .14]	**.12 [.03; 21]**	**.10 [.01; .20]**	.05 [-.04; .14]	*.21* [-.02; .43]	-.09 [-.24; .05]	.01 [-.13; .14]	**-.15 [-.26; -.04]**	.05 [-.17; .27]
SI → SE	.07 [-.11; .26]	.03 [-.11; .17]	*.10 [-.01; .22]*	-.03 [-.11; .05]	*.12 [.00; .24]*	.08 [-.10; .27]	.00 [-.12; .13]	*.10 [-.02; .21]*	-.09 [-.19; .02]	.11 [-.06; .29]
TA*SI → SE	.05 [-.02; .12]	.01 [-.08; .09]	.06 [-.02; .14]	-.04 [-.10; .03]	.05 [-.05; .14]	-.00 [-.05; 04]	-.04 [-.13; .05]	.05 [-.03; .12]	-.05 [-.13; .03]	.01 [-.08; .09]
Family support										
TA→ SE		-.01 [-.11; .09]		.06 [-.04; .15]			-.07 [-.19; .04]		-.10 [-.23; .04]	
FS → SE		**.12 [-.00; .24]**		.04 [-.06; .14]			**.10 [-01; .19]**		.06 [-.03; .14]	
TA*FS → SE		.00 [-.07; .08]		-.02 [-.10; .05]			-.01 [-.09; .08]		.00 [-.07; .08]	

Note. *N*=674; SE = Self-esteem; TA= Teacher attachment (support or rejection); SI = School importance, FS = Family support; In all models, we controlled for prior self-esteem. Bold indicates that the 95% confidence interval did not include zero. Italics indicate that the 90% confidence interval did not include zero.

support from age 11 to 16, positively predicted self-esteem four years post-high school (β = .21). This effect held when controlling for self-esteem at age 11 (β = .17), but not when controlling for self-esteem at age 16. Likewise, a trait factor of teacher rejection negatively predicted self-esteem four years post-high-school (β = -.19). This effect held when controlling for self-esteem at age 11 (β = -.14), but not when controlling for self-esteem at age 16. Thus, the trait factor of teacher attachment predicted self-esteem four years post-high school (1) when *not* controlling for prior self-esteem and (2) when controlling for self-esteem at age 11 (but not when controlling for self-esteem at age 16), which is partially in line with H3.1 and 3.2.

Table 3
Standardized Parameter Estimates [and 95% Confidence Intervals] for Predicting Self-Esteem Four Years Post High School by the Trait Factor of Teacher Support and Rejection

	Teacher support			Teacher rejection		
Predictors	Model 1 Not Controlling for prior SE	Model 2 Controlling for SE at Age 11	Model 3 Controlling for SE at Age 16	Model 1 Not Controlling for prior SE	Model 2 Controlling for SE at Age 11	Model 3 Controlling for SE at Age 16
Trait Factor of TA	**.21** **[.12; .30]**	**.17** **[.05; .28]**	.05 [-.05; .15]	**-.19** **[-.29; -.09]**	**-.14** **[-.26, -.01]**	-.06 [-.17; .04]
Self-esteem Age 11		*.15* *[-.01; .30]*			*.16* *[.01; .32]*	
Self-esteem Age 16			**.48** **[.39; .57]**			**.48** **[.39; .57]**

Note. *N*=674; SE= Self-esteem; TA = Teacher Attachment (support or rejection); Trait factor = Refers to the common variance of teacher support/rejection from age 11 to 16. Bold indicates that the 95% confidence interval did not include zero. Italics indicate that the 90% confidence interval did not include zero.

Invariance across Gender

To test whether the observed effects vary across gender, we repeated all analyses in a multiple-group framework[19]. The results are presented in Table S5, S6, and S7 in the Supplemental Material. We found that, for all research question, results were invariant across gender,

[19] It was not possible to run multiple-group models with latent moderated interactions because the models did not converge. Furthermore, in the preregistration, we also stated that we would run the analyses separated for nativity. For reasons of transparency, we present these results in Table S8, S9, and S10 in the Supplemental material.

because models where coefficients were estimated freely across groups did not show substantial decreases in model fit compared to models where these coefficients were constrained across groups. Despite that, descriptively, we observed that particular relations between self-esteem and later teacher rejection differed in the significance level across groups.

Discussion

The present study investigated the prospective relation between self-esteem and teacher attachment (i.e. teacher support and teacher rejection) in a longitudinal study of 674 Mexican-origin youth from age 11 to 21. Consistent with our preregistered hypotheses, we found positive reciprocal cross-lagged associations between teacher support and self-esteem using both CLPMs and RI-CLPMs. Moreover, also in line with our hypotheses, we found negative reciprocal cross-lagged associations between teacher rejection and self-esteem based on the CLPM, but not based on the RI-CLPM. In contrast to our hypotheses, we did not find that beliefs about school importance or perceived family support moderated associations between teacher support/rejection and self-esteem. Further, in line with our hypotheses, a measure of the stable variance in teacher support (and teacher rejection) from age 11 to 16 positively (negatively) predicted self-esteem four years post-high school (age 21), even when controlling for self-esteem at age 11. However, these prospective associations were not significant when controlling for self-esteem at age 16. The present results were largely invariant across gender.

Teacher Attachment Fuels Students' Sociometer

Several theoretical approaches, such as sociometer theory (Leary & Baumeister, 2000), suggest that healthy self-esteem derives from having positive interpersonal relationships. Numerous studies have documented that individuals who have positive relationships with their parents and peers tend to have higher self-esteem (Harris & Orth, 2019). The present study extends this research by investigating a highly relevant context for social support for adolescents, namely their relationships with teachers. Consistent with theory and research on other social relationships, our findings indicate that teacher attachment predicted year-to-year changes in self-esteem. More specifically, our results suggest that between-person differences (CLPM) in teacher support and rejection prospectively predict individual differences in self-esteem. In other words, students who feel more supported and less rejected by their teachers tend to show subsequent rank-order increases in their self-esteem compared to students who feel less supported and more rejected by their teachers. Consequently, youth who fail to form

healthy attachments with their teachers are at increased risk for the development of low self-esteem compared to those with more positive attachments.

In addition to examining between-person processes, we also examined how within-person fluctuations in teacher attachment predict within-person fluctuations in self-esteem (RI-CLPM). For teacher support, the findings at the within-person level converged with those at the between-person level. Specifically, when students feel more accepted than usual during a particular time period, they will tend to subsequently increase in self-esteem relative to their usual level. This finding suggests that interventions aimed at boosting feelings of teacher support could cause students' to show increases in their self-esteem. In contrast, teacher rejection did not show a within-person effect. That is, students who feel more rejected than usual, do not show a subsequent dip in their self-esteem compared to their usual level. Therefore, our results suggest that there is a closer tie between teacher support and self-esteem than between teacher rejection and self-esteem, which is consistent with previous research indicating that acceptance is more strongly linked to self-esteem than rejection (Blackhart et al., 2009).

In addition to the cross-lagged associations across one year intervals, our results also suggests that teacher attachment during adolescence is linked to self-esteem four years post-high school, even after controlling for initial levels of self-esteem at age 11. This finding emphasizes the long-term importance of teacher attachment on self-esteem and suggest that teachers remain important attachment figures (Kesner, 2000) even when youth are no longer in school. Yet, it is important to note that this prospective effect did not hold when controlling for self-esteem at age 16, suggesting that the self-esteem benefits of healthy teacher attachment persist into young adulthood, but do not increase above and beyond the benefits observed around the end of schooling. Although the majority of the participants in the present study did not go on to higher education, it would be interesting for future research to explore how teacher attachment during the college years relates to self-esteem changes at the between and within-person levels.Overall, our results indicate that teacher attachment robustly predicts self-esteem because this pattern holds across (1) year-to-year between-person effects in CLPMs, (2) year-to-year effects in within-person deviations in RI-CLPMs (only for teacher support), and (3) self-esteem assessed four years post-high school.

In addition to investigating whether teacher attachment predicted students' self-esteem, we analyzed whether two factors moderated this relation. First, we tested an amplification hypothesis that predicts teacher attachment will have a stronger impact on self-esteem for students

who value the importance of school. Contrary to this hypothesis, we did not find evidence of interaction between teacher attachment and school importance. Although this idea was derived theoretically (James, 1890/1963) and is intuitively appealing, the absence of this interaction effect is in line with studies that examined the weighted importance of different self-concept domains in the prediction of global self-esteem, and failed to find unequivocal support (Marsh, 1993, 1995). Moreover, our measure of school importance may have been too broad to capture the value students place on their *social connections* in school (i.e., it assessed how much they value school in general, not teacher attachment in particular).

We also did not find a significant moderating effect of family support, suggesting that previously found interactions between student-parent and student-teacher relations on other student outcomes (e.g., Burchinal et al., 2002; McGrath & van Bergen, 2015; Mitchell-Copeland et al., 1997; Ryan et al., 1994) might not generalize to self-esteem as an outcome. However, in order to investigate the relative importance as well the interactions between different sources of attachment and social support in shaping students' sociometer, future research might benefit from an integrative consideration of different sources of social support (i.e., teachers, parents, peers) using a comparable set of items for each of these sources.

Overall, the present research demonstrates that teacher attachment and self-esteem are prospectively associated from age 11 to 21. In order to understand the mechanisms and processes underlying this relation, future research should embed the present research questions in a more process-oriented design that focuses on when and why students perceive support and rejection from teachers and how they internalize these perceptions and incorporate them into their overall sense of self-worth.

Self-esteem Empowers Positive Student-Teacher Relationships

Along with the positive prospective effects of teacher attachment on self-esteem, we also observed positive prospective effects from self-esteem to teacher attachment. Interestingly, the magnitude of this effect from self-esteem to teacher attachment was consistently stronger than the reverse. A recent meta-analysis examining the association between social relationships and self-esteem indicated that both direction of effects are comparable in size and, thus, the relation between self-esteem and social relationships can be considered a "positive feedback loop" (Harris & Orth, 2019, p.13). The present study further supports this claim and also extends it to the relation between students' self-esteem and their relationship with their teachers, which has not been considered in previous studies. It is important to note that the

effect sizes of all significant cross-paths in the present study were rather small (average β = .08). However, when evaluating the magnitude of the cross-lagged effects, it is important consider that the a) both constructs had high stabilities over time (this already explains a large amount of variance in the outcome variables), (b) self-esteem is determined by a broad variety of factors (e.g. genetics, life events, parenting), and teacher attachment also depends on a variety of factors (e.g., the specific teachers assigned to each student, classroom dynamics, the broader school context, the student and teachers' attachment style, etc.), (c) even small year-to-year effects can produce relatively larger effects across a decade or more of life (e.g., the cumulative impact of teacher attachment from K-12th grade), and (d) the present effect sizes are comparable to a meta-analysis on the relation between social relationships and self-esteem (Harris & Orth, 2019). Therefore, we argue that the present findings are still meaningful and interesting for gaining a deeper understanding of how self-esteem develops during adolescence and beyond.

More generally, our finding that self-esteem predicts teacher attachment provides an entirely new finding for research on student development in school. In particular, self-esteem might be a motor, which helps students to build positive and trustworthy relationships with their teachers. However, based on the present findings, no prediction can be made about whether this effect, instead, follows a risk regulation perspective (Murray et al., 2000; Murray, Holmes, & Collins, 2006). From this perspective, students might internalize their self-esteem and incorporate it into their beliefs about their relationships with others. Alternatively, this effect might be due to a self-broadcasting perspective (Srivastava & Beer, 2005), in which high self-esteem students behave more positively in classroom settings and hence, elicit more support and less rejection from teachers. Future research should explore these processes, for example by using observer ratings of classroom behavior, in order to investigate underlying mechanisms. Relatedly, future research should address whether increases in perceived teacher attachment are related to increases in the perception of other social relationships, such as peer relationships. Finally, it would be informative to understand whether teacher attachment improves from the teacher's as well as the student's perspective.

Limitations

Several limitations should be considered when evaluating the findings. First, the present study relied on self-report data to assess both self-esteem and teacher attachment. Therefore, we cannot not entirely rule out that the results are an artifact due to shared method variance,

such as a general positive self-affirmation bias. However, by including autoregressive terms, we assume that we at least accounted for the major part of this shared method variance (e.g., Orth, Robins, & Roberts, 2008; Sowislo & Orth, 2013). Although self-report is generally considered the gold standard in assessing self-esteem, given its intrinsically phenomenological nature, research on teacher-student relationships sometimes incorporates the perspective of teachers and other informants in addition to the self. Given the specific research questions in the present study, we would consider student self-ratings as the most important source of information because students themselves need to perceive the support/rejection in order to integrate it into their sociometer. However, teacher and/or observer ratings might provide additional important sources of information, in particular when it comes to studying how actual teacher behavior impacts students' perceptions of the quality of their relationship with teachers.

Second, students' reports of teacher rejection were very low at all ages and had relatively little variance, at least in part due to the skewed distribution. We appreciate the fact that students in our sample, for the most part, did not feel very rejected by their teachers, however this also restricted the variance, and in particular the reliable variance (as seen in the relatively low internal consistency values), that could be explained by students' self-esteem. As a possible consequence, the observed associations with teacher rejection were less robust across all modeling procedures and were not invariant across gender. We cannot rule out that our measure of teacher rejection did not capture the full range of individual differences in teacher rejection, for example, because students answered the questions in a socially desirable way. Accordingly, future research should attempt to replicate the present findings using a more reliable measure of perceived teacher rejection.

Third, given the demographic make-up of our sample (i.e., Mexican-origin adolescents), we cannot make broad claims about the generalizability of the present findings. Even though we have no reason to expect that the overall pattern of relations between self-esteem and teacher attachment will be different in other populations, it is an empirical question whether the findings and effect sizes generalize to other samples. For example, it is possible that ethnic minority students, who are at particular risk for adverse school outcomes, benefit more from teacher attachment than youth from majority racial/ethnic groups (Garcia-Reid et al., 2005; Roorda et al., 2011). Therefore, this research should be replicated in other diverse samples of students.

Conclusion

The present study provides longitudinal support for reciprocal relations between teacher attachment (i.e., teacher support and teacher rejection) and self-esteem from age 11 to 21. Not only does teacher attachment prospectively predict students' self-esteem, but also, and to an even greater extent, self-esteem predicts positive perceptions of teacher attachment (especially teacher support). Thus, the relation between teacher attachment and self-esteem should be viewed as a dynamic transactional process that involves reciprocal associations over time. Finally, the present research has an important message for theories of self-esteem development. Cooley (1902) noted more than a century ago that the perceptions of "significant others" serve as a "looking glass" reflecting back our worth as a person. The present findings demonstrate that teachers are an important source of these reflections and should be accorded the status of "significant other".

References

Aldrup, K., Klusmann, U., Lüdtke, O., Göllner, R., & Trautwein, U. (2018). Social support and classroom management are related to secondary students' general school adjustment: A multilevel structural equation model using student and teacher ratings. *Journal of Educational Psychology*, *110*(8), 1066–1083. https://doi.org/10.1037/edu0000256

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, *112*(4), 545–557. https://doi.org/10.1037/0021-843X.112.4.545

Armsden, G. C., & Greenberg, M. T. (1987). The inventory of parent and peer attachment: Individual differences and their relationship to psychological well-being in adolescence. *Journal of Youth and Adolescence*, *16*(5), 427–454. https://doi.org/10.1007/BF02202939

Arnett, J. J. (1999). Adolescent storm and stress, reconsidered. *American Psychologist*, *54*(5), 317–326. https://doi.org/10.1037//0003-066x.54.5.317

Barber, B. K., Chadwick, B. A., & Oerter, R. (1992). Parental behaviors and adolescent self-esteem in the United States and Germany. *Journal of Marriage and the Family*, *54*(1), 128. https://doi.org/10.2307/353281

Blackhart, G. C., Nelson, B. C., Knowles, M. L., & Baumeister, R. F. (2009). Rejection elicits emotional reactions but neither causes immediate distress nor lowers self-esteem: a meta-analytic review of 192 studies on social exclusion. *Personality and Social Psychology Review*, *13*(4), 269–309. https://doi.org/10.1177/1088868309346065

Block, J., & Robins, R. W. (1993). A longitudinal study of consistency and change in self-esteem from early adolescence to early adulthood. *Child Development*, *64*(3), 909. https://doi.org/10.2307/1131226

Boudreault-Bouchard, A.-M., Dion, J., Hains, J., Vandermeerschen, J., Laberge, L., & Perron, M. (2013). Impact of parental emotional support and coercive control on adolescents' self-esteem and psychological distress: results of a four-year longitudinal study. *Journal of Adolescence*, *36*(4), 695–704. https://doi.org/10.1016/j.adolescence.2013.05.002

Bowlby, J. (1969). *Attachment and loss: Vol. 1. Attachment*. New York, NY: Basic Books.

Bowlby, J. (1973). *Attachment and loss: Vol. 2. Seperation: Anxiety and anger*. New York, NY: Basic Books.

Bowlby, J. (1980). *Attachment and loss: Vol. 3. Loss: Sadness and depression*. New York, NY: Basic Books.

Brey, C. de, Musu, L., McFarland, J., Wilkinson-Flicker, S., Diliberti, M., Zhang, A.,. . . Wang, X. (2019). Status and Trends in the Education of Racial and Ethnic Groups 2018 (NCES 2019-038). Retrieved from https://nces.ed.gov/pubsearch/

Burchinal, M. R., Peisner-Feinberg, E., Pianta, R., & Howes, C. (2002). Development of academic skills from preschool through second grade: Family and classroom predictors of developmental trajectories. *Journal of School Psychology*, *40*(5), 415–436. https://doi.org/10.1016/S0022-4405(02)00107-3

Castro-Schilo, L., Ferrer, E., Hernández, M. M., & Conger, R. D. (2016). Developmental outcomes of school attachment among students of Mexican origin. *Journal of Research on Adolescence*, *26*(4), 753–768. https://doi.org/10.1111/jora.12223

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, *12*(4), 381–398. https://doi.org/10.1037/1082-989X.12.4.381

Cooley, C. H. (1902). *Human nature and the social order*. New York: Charles Scribner´s Sons.

Creemers, B., Kyriakides, L., & Antoniou, P. (2013). Establishing theoretical frameworks to describe teacher effectiveness. In B. P. M. Creemers, L. Kyriakidēs, & P. Antoniou (Eds.), *Teacher professional development for improving quality of teaching* (pp. 101–135). Dordrecht, New York: Springer. https://doi.org/10.1007/978-94-007-5207-8_7

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Springer Science+Business Media.

Donnellan, M. B., Trzesniewski, K. H., & Robins, R. W. (2011). Self-esteem: Enduring issues and controversies. In T. Chamorro-Premuzic (Ed.), *The Wiley-Blackwell handbook of individual differences* (pp. 718–746). Chichester: Wiley-Blackwell.

Eccles, J., & Midgley, C. (1989). Stage-environment fit: Developmentally appropriate classrooms for young adolescents. In R. Ames & C. Ames (Eds.), *Research on motivation in education: Vol. 3. Goals and cognitions* (pp. 139–186). New York, NY: Academic Press.

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Mac Iver, D. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist*, *48*(2), 90–101. https://doi.org/10.1037/0003-066X.48.2.90

Eccles, J. S., & Roeser, R. W. (2011). Schools as developmental contexts during adolescence. *Journal of Research on Adolescence*, *21*(1), 225–241. https://doi.org/10.1111/j.1532-7795.2010.00725.x

Erikson, E. H. (1968). *Identity, Youth and Crisis*. New York: Norton.

Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*, *95*(1), 148–162. https://doi.org/10.1037/0022-0663.95.1.148

Garcia-Reid, P., Reid, R. J., & Peterson, N. A. (2005). School engagement among latino youth in an urban middle school context. *Education and Urban Society*, *37*(3), 257–275. https://doi.org/10.1177/0013124505275534

Gruenenfelder-Steiger, A. E., Harris, M. A., & Fend, H. A. (2016). Subjective and objective peer approval evaluations and self-esteem development: A test of reciprocal, prospective, and long-term effects. *Developmental Psychology*, *52*(10), 1563–1577. https://doi.org/10.1037/dev0000147

Hall, G. S. (1904). *Adolescence: Its psychology and its relation to physiology, anthropology, sociology, sex, crime, religion, and education (Vols. I & II)*. Englewood Cliffs, NJ: Prentice-Hall.

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

Hamre, B. K., & Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*, *72*(2), 625–638. https://doi.org/10.1111/1467-8624.00301

Harris, M. A., Gruenenfelder-Steiger, A. E., Ferrer, E., Donnellan, M. B., Allemand, M., Fend, H.,. . . Trzesniewski, K. H. (2015). Do parents foster self-esteem? Testing the prospective

impact of parent closeness on adolescent self-esteem. *Child Development*, *86*(4), 995–1013. https://doi.org/10.1111/cdev.12356

Harris, M. A., & Orth, U. (2019). The link between self-esteem and social relationships: A meta-analysis of longitudinal studies. *Journal of Personality and Social Psychology*. Advance online publication. https://doi.org/10.1037/pspp0000265

Helmke, A. (2012). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts* (4., aktualisierte Aufl.). *Schule weiterentwickeln, Unterricht verbessern Orientierungsband*. Seelze-Velber: Klett/Kallmeyer. Retrieved from http://www.socialnet.de/rezensionen/isbn.php?isbn=978-3-7800-1009-4

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29. https://doi.org/10.1038/466029a

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

Hughes, J. N., & Cao, Q. (2018). Trajectories of teacher-student warmth and conflict at the transition to middle school: Effects on academic engagement and achievement. *Journal of School Psychology*, *67*, 148–162. https://doi.org/10.1016/j.jsp.2017.10.003

James, W. (1890/1963). *The principles of psychology*. New York: Holt, Rinehart & Winston.

Kesner, J. E. (2000). Teacher characteristics and the quality of child–teacher relationships. *Journal of School Psychology*, *38*(2), 133–149. https://doi.org/10.1016/S0022-4405(99)00043-6

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*(4), 457–474. https://doi.org/10.1007/BF02296338

Krauss, S., Orth, U., & Robins, R. W. (2019). Family environment and self-esteem development: A longitudinal study from age 10 to 16. *Journal of Personality and Social Psychology.* Advance online publication. https://doi.org/10.1037/pspp0000263

Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, *9*(3), 231–251. https://doi.org/10.1007/s10984-006-9015-7

Kunter, M., & Voss, T. (2013). The model of instructional quality in COACTIV: A multicriteria analysis. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers. Mathematics teacher education, Vol. 8* (pp. 97–124). New York, London: Springer. https://doi.org/10.1007/978-1-4614-5149-5_6

Leary, M. R., & Baumeister, R. F. (2000). The nature and function of self-esteem: Sociometer theory. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 1–62). San Diego, CA: Academic Press. https://doi.org/10.1016/S0065-2601(00)80003-9

Leary, M. R., Tambor, E. S., Terdal, S. K., & Downs, D. L. (1995). Self-esteem as an interpersonal monitor: The sociometer hypothesis. *Journal of Personality and Social Psychology*, *68*(3), 518–530. https://doi.org/10.1037/0022-3514.68.3.518

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 151–173. https://doi.org/10.1207/S15328007SEM0902_1

Lynch, M., & Cicchetti, D. (1997). Children's relationships with adults and peers: An examination of elementary and junior high school students. *Journal of School Psychology*, *35*(1), 81–99. https://doi.org/10.1016/S0022-4405(96)00031-3

Marsh, H. W. (1993). Relations between global and specific domains of self: The importance of individual importance, certainty, and ideals. *Journal of Personality and Social Psychology*, *65*(5), 975–992. https://doi.org/10.1037/0022-3514.65.5.975

Marsh, H. W. (1995). A Jamesian model of self-investment and self-esteem: Comment on Pelham (1995). *Journal of Personality and Social Psychology*, *69*(6), 1151–1160. https://doi.org/10.1037/0022-3514.69.6.1151

Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J. S., & Davier, M. von. (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*(3), 257–284. https://doi.org/10.1037/a0032773

Martin, A. J., & Collie, R. J. (2019). Teacher–student relationships and students' engagement in high school: Does the number of negative and positive relationships with teachers matter? *Journal of Educational Psychology*, *111*(5), 861–876. https://doi.org/10.1037/edu0000317

McGrath, K. F., & van Bergen, P. (2015). Who, when, why and to what end? Students at risk of negative student–teacher relationships and their outcomes. *Educational Research Review, 14*, 1–17. https://doi.org/10.1016/j.edurev.2014.12.001

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *The Journal of Applied Psychology*, *93*(3), 568–592. https://doi.org/10.1037/0021-9010.93.3.568

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. https://doi.org/10.1007/BF02294825

Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). Change in teacher efficacy and student self- and task-related beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology*, *81*(2), 247–258. https://doi.org/10.1037/0022-0663.81.2.247

Mitchell-Copeland, J., Denham, S. A., & DeMulder, E. K. (1997). Q-sort assessment of child–teacher attachment relationships and social competence in the preschool. *Early Education & Development*, *8*(1), 27–39. https://doi.org/10.1207/s15566935eed0801_3

Murray, S. L., Holmes, J. G., & Collins, N. L. (2006). Optimizing assurance: the risk regulation system in relationships. *Psychological Bulletin*, *132*(5), 641–666. https://doi.org/10.1037/0033-2909.132.5.641

Murray, S. L., Holmes, J. G., & Griffin, D. W. (2000). Self-esteem and the quest for felt security: How perceived regard regulates attachment processes. *Journal of Personality and Social Psychology*, *78*(3), 478–498. https://doi.org/10.1037//0022-3514.78.3.478

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide (8th ed.)*. Los Angeles, CA: Muthén & Muthén.

Nezlek, J. B., Kowalski, R. M., Leary, M. R., Blevins, T., & Holgate, S. (1997). Personality moderators of reactions to interpersonal rejection: Depression and trait Self-Esteem. *Personality and Social Psychology Bulletin*, *23*(12), 1235–1244. https://doi.org/10.1177/01461672972312001

OECD. (2016). *Immigrant background, student performance and students' attitudes towards science*. Paris: PISA 2015 Results (Volume I): Excellence and Equity in Education, OECD Publishing.

OECD. (2018). *The Resilience of Students with an Immigrant Background*. Paris: OECD Reviews of Migrant Education, OECD Publishing.

Orth, U., Clark, D. A., Donnellan, M. B., & Robins, R. W. (in press). Testing Prospective Effects in Longitudinal Research: Comparing Seven Competing Cross-Lagged Models. *Journal of Personality and Social Psychology*.

Orth, U., & Robins, R. W. (2019). Development of self-esteem across the lifespan. In D. P. McAdams, R. L. Shiner, & J. L. Tackett (Eds.), *Handbook of personality development* (pp. 328–344). New York, NY: The Guilford Press.

Orth, U. (2018). The family environment in early childhood has a long-term effect on self-esteem: A longitudinal study from birth to age 27 years. *Journal of Personality and Social Psychology*, *114*(4), 637–655. https://doi.org/10.1037/pspp0000143

Orth, U., Erol, R. Y., & Luciano, E. C. (2018). Development of self-esteem from age 4 to 94 years: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *144*(10), 1045–1080. https://doi.org/10.1037/bul0000161

Orth, U., Robins, R. W., & Roberts, B. W. (2008). Low self-esteem prospectively predicts depression in adolescence and young adulthood. *Journal of Personality and Social Psychology*, *95*(3), 695–708. https://doi.org/10.1037/0022-3514.95.3.695

Orth, U., Robins, R. W., & Widaman, K. F. (2012). Life-span development of self-esteem and its effects on important life outcomes. *Journal of Personality and Social Psychology*, *102*(6), 1271–1288. https://doi.org/10.1037/a0025558

Pianta, R. C., & Steinberg, M. (1992). Teacher-child relationships and the process of adjusting to school. *New Directions for Child and Adolescent Development*, *1992*(57), 61–80. https://doi.org/10.1002/cd.23219925706

Pössel, P., Rudasill, K. M., Sawyer, M. G., Spence, S. H., & Bjerg, A. C. (2013). Associations between teacher emotional support and depressive symptoms in Australian adolescents: a 5-year longitudinal study. *Developmental Psychology*, *49*(11), 2135–2146. https://doi.org/10.1037/a0031767

Reddy, R., Rhodes, J. E., & Mulhall, P. (2003). The influence of teacher support on student adjustment in the middle school years: A latent growth curve study. *Development and Psychopathology*, *15*(1), 119–138. https://doi.org/10.1017/s0954579403000075

Reitz, A. K., Motti-Stefanidi, F., & Asendorpf, J. B. (2016). Me, us, and them: Testing sociometer theory in a socially diverse real-life context. *Journal of Personality and Social Psychology*, *110*(6), 908–920. https://doi.org/10.1037/pspp0000073

Rice, L., Barth, J. M., Guadagno, R. E., Smith, G. P. A., McCallum, D. M., & ASERT. (2013). The role of social support in students' perceived abilities and attitudes toward math and science. *Journal of Youth and Adolescence, 42*(7), 1028–1040. https://doi.org/10.1007/s10964-012-9801-8

Roberts, B. W., & Wood, D. (2006). Personality development in the context of the neo-socio-analytic model of personality. In D. Mroczek & T. Little (Eds.), *Handbook of personality development.* Mahwah, NJ: Erlbaum.

Robins, R. W., Tracy, J. L., Trzesniewski, K., Potter, J., & Gosling, S. D. (2001). Personality correlates of self-esteem. *Journal of Research in Personality, 35*(4), 463–482. https://doi.org/10.1006/jrpe.2001.2324

Roeser, R. W., Lord, S. E., & Eccles, J. (1994). *A portrait of academic alienation in adolescence: Motivation, mental health, and family experience.* Paper presented at the Biennial Meeting of the Society for Research on Adolescence, San Diego, CA.

Roorda, D. L., Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher–student relationships on students' school engagement and achievement. *Review of Educational Research, 81*(4), 493–529. https://doi.org/10.3102/0034654311421793

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rosenberg, M., Schooler, C., & Schoenbach, C. (1989). Self-esteem and adolescent problems: Modeling reciprocal effects. *American Sociological Review, 54*(6), 1004. https://doi.org/10.2307/2095720

Ruzek, E. A., Hafen, C. A., Allen, J. P., Gregory, A., Mikami, A. Y., & Pianta, R. C. (2016). How teacher emotional support motivates students: The mediating roles of perceived peer relatedness, autonomy support, and competence. *Learning and Instruction, 42*, 95–103. https://doi.org/10.1016/j.learninstruc.2016.01.004

Ryan, R. M., Stiller, J. D., & Lynch, J. H. (1994). Representations of relationships to teachers, parents, and friends as predictors of academic motivation and self-esteem. *The Journal of Early Adolescence, 14*(2), 226–249. https://doi.org/10.1177/027243169401400207

Sameroff, A. J. (2009). *The transactional model of development: How children and contexts shape each other* (1st ed.). Washington, D.C: American Psychological Association. Retrieved from http://search.ebscohost.com/direct.asp?db=pzh&jid=%22200905801%22&scope=site

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness* (1. ed.). Oxford: Pergamon.

Scherer, R., Nilsen, T., & Jansen, M. (2016). Evaluating individual students' perceptions of instructional quality: An investigation of their factor structure, measurement invariance, and relations to educational outcomes. *Frontiers in Psychology*, *7*, 110. https://doi.org/10.3389/fpsyg.2016.00110

Scherrer, V., & Preckel, F. (2019). Development of motivational variables and self-esteem during the school career: A meta-analysis of longitudinal studies. *Review of Educational Research*, *89*(2), 211–258. https://doi.org/10.3102/0034654318819127

Smith, E. P., Connell, C. M., Wright, G., Sizer, M., Norman, J. M., Hurley, A., & Walker, S. N. (1997). An ecological model of home, school, and community partnerships: Implications for research and practice. *Journal of Educational and Psychological Consultation*, *8*(4), 339–360. https://doi.org/10.1207/s1532768xjepc0804_2

Soto, C. J., & Tackett, J. L. (2015). Personality traits in childhood and adolescence. *Current Directions in Psychological Science*, *24*(5), 358–362. https://doi.org/10.1177/0963721415589345

Sowislo, J. F., & Orth, U. (2013). Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies. *Psychological Bulletin*, *139*(1), 213–240. https://doi.org/10.1037/a0028931

Srivastava, S., & Beer, J. S. (2005). How self-evaluations relate to being liked by others: Integrating sociometer and attachment perspectives. *Journal of Personality and Social Psychology*, *89*(6), 966–977. https://doi.org/10.1037/0022-3514.89.6.966

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality*, *13*(5), 389–408. https://doi.org/10.1002/(SICI)1099-0984(199909/10)13:5<389::AID-PER361>3.0.CO;2-A

Swann, W. B., JR, Chang-Schneider, C., & Larsen McClarty, K. (2007). Do people's self-views matter? Self-concept and self-esteem in everyday life. *American Psychologist*, *62*(2), 84–94. https://doi.org/10.1037/0003-066X.62.2.84

Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2003). Stability of self-esteem across the life span. *Journal of Personality and Social Psychology*, *84*(1), 205–220.

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

Verschueren, K., Marcoen, A., & Schoefs, V. (1996). The internal working model of the self, attachment, and competence in five-year-olds. *Child Development*, *67*(5), 2493–2511. https://doi.org/10.1111/j.1467-8624.1996.tb01870.x

Wagner, J., Lüdtke, O., Robitzsch, A., Göllner, R., & Trautwein, U. (2018). Self-esteem development in the school context: The roles of intrapersonal and interpersonal social predictors. *Journal of Personality*, *86*(3), 481–497. https://doi.org/10.1111/jopy.12330

Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B., & Trautwein, U. (2016). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology*, *108*(5), 705–721. https://doi.org/10.1037/edu0000075

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037//0022-3514.54.6.1063

Wentzel, K. R. (2009). Students' relationship with teachers as motivational contexts. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 301–322). Mahwah, NJ: Taylor and Francis.

Whitbeck, L. B., Simons, R. L., Conger, R. D., Lorenz, F. O., Huck, S., & Elder, G. H. (1991). Family economic hardship, parental support, and adolescent self-esteem. *Social Psychology Quarterly*, *54*(4), 353. https://doi.org/10.2307/2786847

Wigfield, A., & Eccles, J. S. (1994). Children's competence beliefs, achievement values, and general self-esteem. *The Journal of Early Adolescence*, *14*(2), 107–138. https://doi.org/10.1177/027243169401400203

Wigfield, A., Eccles, J. S., Mac Iver, D., Reuman, D. A., & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology, 27*(4), 552–565. https://doi.org/10.1037/0012-1649.27.4.552

Zimet, G. D., Dahlem, N. W., Zimet, S. G., & Farley, G. K. (1988). The multidimensional scale of perceived social support. *Journal of Personality Assessment, 52*(1), 30–41. https://doi.org/10.1207/s15327752jpa5201_2

Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z.,. . . Diener, E. (2019). From data to causes I: Building a general cross-lagged panel model (GCLM). *Organizational Research Methods, 175*(4), 109442811984727. https://doi.org/10.1177/1094428119847278

Supplemental Material

Table S1
Attrition Analysis: Differences in Relevant Outcomes Variables

Dependent variable	*N*	*d*	*p*
Teacher Support T1	567	0.158	.083
Teacher Support T2	578	0.136	.129
Teacher Rejection T1	567	-0.049	.590
Teacher Rejection T2	578	0.029	.745
Self-Esteem T1	309	-0.027	.826
Self-Esteem T2	576	0.153	.086
School Importance T1	567	0.129	.157
School Importance T2	578	0.020	.821
Gender (0 = female, 1 = male)	674	*OR* = 0.764	.085

Note. N=674; d = standardized mean difference; independent variable = 1 indicates missing at a minimum of one time point; higher GPA indicates lower achievement.

Table S2

Latent Means and Standard Deviations, Internal Consistencies and Rank Order Stabilities

	Self-esteem			Teacher support			Teacher rejection			School Importance			Family Support		
Age	*M(SD)*	ω	$r_{t,t-1}$	*M(SD)*	ω	$r_{t,t-1}$	*M(SD)*	ω	$r_{t,t-1}$	*M(SD)*	ω	$r_{t,t-1}$	*M(SD)*	ω	$r_{t,t-1}$
11	3.27 (0.39)	.80	-	3.04 (0.56)	.85	-	1.11 (0.19)	.58	-	3.70 (0.32)	.71	-			-
12	3.19 (0.38)	.78	.55	2.68 (0.70)	.90	.56	1.16 (0.24)	.62	.40	3.60 (0.34)	.71	.49	3.37 (0.58)	.88	
13	3.25 (0.41)	.83	.68	2.51 (0.73)	.91	.56	1.11 (0.22)	.71	.36	3.56 (0.38)	.78	.50			
14	3.13 (0.40)	.86	.56	2.48 (0.77)	.93	.60	1.24 (0.20)	.63	.45	3.46 (0.43)	.80	.55	3.16 (0.68)	.91	.53
15	3.18 (0.42)	.86	.70	2.46 (0.74)	.92	.52	1.12 (0.22)	.69	.68	3.47 (0.38)	.79	.60			
16	3.10 (0.40)	.86	.61	2.43 (0.73)	.92	.55	1.11 (0.20)	.69	.61	3.41 (0.42)	.79	.63	3.02 (0.72)	.91	.56
17	3.23 (0.43)	.88	.72	-	-	-	-	-	-	-	-	-	-	-	-
18	3.19 (0.39)	.87	.63	-	-	-	-	-	-	-	-	-	-	-	-
19	3.12 (0.45)	.90	.55	-	-	-	-	-	-	-	-	-	-	-	-
21	3.03 (0.37)	.86	.53	-	-	-	-	-	-	-	-	-	-	-	-

Note. N=674.

Table S3
Measurement Invariance over Time Separately for Each Instrument

	χ^2	*df*	SF	CFI	TLI	RMSEA	SRMR	$\Delta\chi^2$	Δdf	Δp
Self-esteem										
1a. Free loadings	258.62	225	1.075	0.997	0.993	0.015	0.026			
2a. Constraints on loadings	283.07	243	1.077	0.996	0.993	0.016	0.038	24.35	18	0.144
3a. Constraints on loadings and intercepts	406.52	261	1.070	0.985	0.975	0.029	0.037	133.69	18	0.000
Teacher support										
1b. Free loadings	76.485	75	1.063	1.000	1.000	0.005	0.018			
2b. Constraints on loadings	104.20	85	1.036	0.998	0.996	0.019	0.029	31.99	10	0.000
3c. Constraints on loadings and intercepts	126.93	95	1.032	0.996	0.994	0.023	0.033	15.93	10	0.102
Teacher rejection										
1c. Free loadings	211.49	192	1.595	0.982	0.975	0.012	0.039			
2c. Constraints on loadings	214.76	207	1.748	0.993	0.991	0.008	0.043	10.27	15	0.802
3c. Constraints on loadings and intercepts	231.00	222	1.694	0.992	0.990	0.008	0.043	16.77	15	0.333
School importance										
1d. Free loadings	95.06	75	1.292	0.994	0.988	0.020	0.039			
2d. Constraints on loadings	104.22	85	1.356	0.994	0.989	0.019	0.052	10.08	10	0.433
3d. Constraints on loadings and intercepts	157.67	95	1.318	0.981	0.969	0.032	0.064	66.58	10	0.000
Family support										
1e. Free loadings	110.52	39	1.148	0.979	0.965	0.054	0.025			
2e. Constraints on loadings	117.38	45	1.126	0.979	0.970	0.050	0.030	5.36	6	0.498
3e. Constraints on loadings and intercepts	123.52	51	1.111	0.979	0.973	0.047	0.033	5.11	6	0.530

Note. N=674

Table S4
Fit Indices for the Tested Structural Models

	χ^2	*df*	SF	CFI	TLI	RMSEA	SRMR	$\Delta\chi^2$	Δdf	Δp
Self-Esteem & Teacher Support										
Model TS1 (CLPM, free across time)	954.25	522	1.046	0.968	0.961	0.036	0.079			
Model TS2 (CLPM, constrained across time)	994.62	538	1.047	0.966	0.960	0.036	0.086	40.20	16	0.000
Model TS3a (CLPM, gender multiple group, free[a])	1671.37	1076	1.025	0.957	0.950	0.041	0.095			
Model TS3b (CLPM, gender multiple group, constrained[b])	1674.81	1080	1.026	0.957	0.950	0.041	0.094	4.02	4	0.404
Model TS4 (RI-CLPM, free across time)	773.84	515	1.044	0.981	0.977	0.028	0.050			
Model TS5 (RI-CLPM, constrained time)	810.74	531	1.046	0.979	0.975	0.028	0.056	36.90	16	0.003
Model TS6a (RI-CLPM, gender multiple group, free[a])	1510.70	1074	1.025	0.969	0.963	0.035	0.075			
Model TS6b (RI-CLPM, gender multiple group, constrained[b])	1559.41	1078	1.103	0.965	0.959	0.037	0.075	7.47	4	0.113
Model TS7 (LST Model)	351.62	162	1.050	0.979	0.972	0.042	0.094			
Model TS8 (LST Model gender multiple group)	506.58	332	1.043	0.980	0.975	0.040	0.100			
Self-Esteem & Teacher Rejection										
Model TR1 (CLPM, free across time)	1063.26	757	1.206	0.948	0.941	0.025	0.057			
Model TR2 (CLPM, constrained across time)	1108.78	773	1.224	0.943	0.937	0.026	0.065	36.10	16	0.003
Model TR3a (CLPM, gender multiple group, free[a])	2352.13	1546	1.103	0.882	0.868	0.040	0.088			
Model TR3b (CLPM, gender multiple group, constrained[b])	2362.75	1550	1.108	0.881	0.867	0.040	0.091	7.78	4	0.100
Model TR4 (RI-CLPM, free across time)	973.61	750	1.198	0.962	0.956	0.021	0.050			
Model TR5 (RI-CLPM, constrained across time)	1017.81	766	1.215	0.957	0.952	0.022	0.057	44.20	16	0.004
Model TR6a (RI-CLPM, gender multiple group, free[a])	2287.07	1542	1.090	0.891	0.878	0.038	0.080			
Model TR6b (RI-CLPM, gender multiple group, constrained[b])	2303.05	1546	1.101	0.889	0.876	0.039	0.083	8.02	4	0.091
Model TR7 (LST Model)	370.98	307	1.554	0.963	0.958	0.018	0.060			
Model TR8 (LST Model, gender multiple group)	883.14	624	1.393	0.872	0.856	0.035	0.092			

Note. N=674; SE = Self-esteem; TS = Teacher support; TR= Teacher rejection
[a]Refers to freely estimated cross-lagged and autoregressive coefficients across groups; Models include constraints across time
[b]Refers to constraints on the cross-lagged and autoregressive coefficients across time and across groups

Table S5

Standardized Parameter Estimates [and 95% Confidence Intervals] of the (Random Intercept) Cross-Lagged Panel Model (RI-CLPM; CLPM) between Self-Esteem and Teacher Support

	CLPM			RI-CLPM		
Path	All, freely estimated	Female, longitudinal. constraints	Male, longitudinal constraints	All, freely estimated	Female, longitudinal. constraints	Male, longitudinal constraints
Autoregressive						
TS11→TS12	**.52 [.43; .60]**	**.46 [.40; .53]**	**.48 [.40; .56]**	**.24 [.10; .38]**	**.40 [.31; .49]**	**.21 [.11; .31]**
TS12→TS13	**.54 [.47; .61]**	**.50 [.43; .57]**	**.55 [.47; .62]**	**.28 [.16; .40]**	**.44 [.34; .53]**	**.26 [.14; .38]**
TS13→TS14	**.60 [.54; .67]**	**.53 [.45; .60]**	**.55 [.47; .63]**	**.41 [.30; .52]**	**.47 [.36; .58]**	**.25 [.12; .38]**
TS14→TS15	**.51 [.43; .59]**	**.51 [.43; .58]**	**.57 [.49; .65]**	**.27 [.15; .40]**	**.43 [.32; .54]**	**.29 [.15; .42]**
TS15→TS16	**.53 [.45; .60]**	**.58 [.51; .65]**	**.56 [.49; .64]**	**.31 [.19; .42]**	**.52 [.42; .62]**	**.26 [.13; .39]**
SE11→SE12	**.61 [.49; .74]**	**.63 [.56; .70]**	**.59 [.51; .67]**	**.29 [.07; .51]**	**.54 [.40; .68]**	*.17 [.01; .34]*
SE12→SE13	**.71 [.63; .78]**	**.68 [.59; .77]**	**.65 [.56; .75]**	**.43 [.26; .60]**	**.58 [.42; .75]**	*.15 [-.01; .31]*
SE13→SE14	**.56 [.48; .65]**	**.61 [.54; .68]**	**.61 [.51; .70]**	**.21 [.02; .39]**	**.52 [.38; .66]**	*.16 [.01; .31]*
SE14→SE15	**.70 [.64; .76]**	**.64 [.55; .72]**	**.69 [.60; .77]**	**.47 [.34; .60]**	**.53 [.37; .69]**	*.19 [-.01; .39]*
SE15→SE16	**.60 [.52; .68]**	**.66 [.59; .73]**	**.56 [.48; .65]**	**.34 [.20; .48]**	**.57 [.43; .71]**	*.12 [-.01; .25]*
Cross-lagged						
SE11→TS12	**.18 [.06; .29]**	**.17 [.11; .23]**	**.10 [.04; .16]**	.11 [-.06; .28]	**.28 [.18; .38]**	.08 [-.05; .21]
SE12→TS13	**.17 [.08; .26]**	**.16 [.10; .22]**	**.10 [.04; .17]**	**.15 [.01; .29]**	**.26 [.17; .36]**	.07 [-.04; .18]
SE13→TS14	.05 [-.04; .14]	**.15 [.10; .20]**	**.09 [.04; .15]**	.00 [-.14; .13]	**.25 [.16; .34]**	.07 [-.04; .17]
SE14→TS15	.06 [-.02; .15]	**.15 [.10; .20]**	**.10 [.04; .15]**	.08 [-.05; .21]	**.24 [.15; .32]**	.07 [-.05; .19]
SE15→TS16	**.15 [.06; .23]**	**.16 [.10; .22]**	**.09 [.03; .14]**	*.12 [.00; .26]*	**.27 [.17; .37]**	.06 [-.04; .15]
TS11→SE12	-.02[-.13; .09]	**.08 [.04; .12]**	**.04 [.00; .08]**	*-.15[-.31; .01]*	**.16 [.09; .22]**	*.07 [.00; .14]*
TS12→SE13	.05 [-.03; .14]	**.10 [.05; .14]**	**.05 [.00; .10]**	.02 [-.11; .14]	**.20 [.12; .28]**	*.09 [.00; .18]*
TS13→SE14	**.14 [.06; .22]**	**.10 [.05; .15]**	**.05 [.01; .10]**	**.17 [.03; .31]**	**.20 [.12; .29]**	*.09 [.00; .19]*
TS14→SE15	.06 [-.01; .14]	**.10 [.05; .15]**	**.06 [.01; .12]**	.06 [-.07; .19]	**.20 [.12; .28]**	*.12 [.00; .25]*
TS15→SE16	*.07 [-.01; .16]*	**.11 [.06; .16]**	**.06 [.00; .11]**	*.10 [-.02; .22]*	**.23 [.14; .32]**	*.09 [.00; .18]*

Note. *N*=674; SE = Self-esteem; TS= Teacher support. Bold indicates that the 95% confidence interval did not include zero. Italics indicate that the 90% confidence interval did not include zero.

Table S6

Standardized Parameter Estimates [and 95% Confidence Intervals] of the Cross-Lagged Panel Model (CLPM) between Self-Esteem and Teacher Rejection

	CLPM			RI-CLPM		
Path	All, freely estimated	Female, longitudinal. constraints	Male, longitudinal constraints	All, freely estimated	Female, longitudinal. constraints	Male, longitudinal constraints
Autoregressive						
TR11→TR12	**.30 [.05; .55]**	**.36 [.12; .61]**	**.56 [.36; .77]**	.03 [-.34; .40]	.13[-.15; .40]	**.67 [.33; 1.01]**
TR12→TR13	**.44 [.24; .64]**	**.50 [.28; .71]**	**.59 [.40; .78]**	.16 [-.20; .52]	.19[-.27; .65]	**.63 [.37; .89]**
TR13→TR14	**.65[.45; .85]**	**.52 [.29; .74]**	**.70 [.46; .94]**	.39[-.25; 1.02]	.22[-.21; .64]	**.77 [.54; 1.00]**
TR14→TR15	**.83[.61; 1.05]**	**.56 [.22; .90]**	**.66 [.45; .87]**	.68[.12; 1.24]	.20[-.34; .75]	**.74 [.50; .97]**
TR15→TR16	**.62 [.47; .77]**	**.47 [.20; .73]**	**.68 [.54; .83]**	**.46 [.18; .74]**	.20[-.22; .62]	**.74 [.53; .95]**
SE11→SE12	**.66 [.52; .79]**	**.63 [.55; .70]**	**.58 [.50; .66]**	**.35 [.09; .62]**	.51[-2.27; 3.29]	**.24 [.06; .42]**
SE12→SE13	**.69 [.61; .77]**	**.71 [.62; .80]**	**.64 [.55; .74]**	**.42 [.26; .59]**	.69[.29; 1.10]	**.19 [.02; .37]**
SE13→SE14	**.59 [.51; .68]**	**.62 [.55; .69]**	**.61 [.51; .70]**	**.26 [.07; .44]**	.63[-.56; 1.83]	**.20 [.04; .36]**
SE14→SE15	**.70 [.63; .76]**	**.66 [.57; .74]**	**.69 [.60; .77]**	**.49 [.36; .63]**	.69[-.83; 2.20]	**.24 [.02; .47]**
SE15→SE16	**.62 [.54; .69]**	**.68 [.61; .74]**	**.57 [.48; .65]**	**.37 [.23; .51]**	.70[-.74; 2.14]	**.15 [.02; .29]**
Cross-lagged						
SE11→TR12	**-.30[-.47; -.14]**	**-.06[-.12; -.01]**	**-.08[-.15; -.02]**	**-.32[-.62; -.02]**	-.05 [-.77; .66]	**.17 [.02; .33]**
SE12→TR13	-.07[-.19; .05]	**-.08[-.15; -.01]**	**-.08[-.15; -.01]**	.05[-.19; .29]	-.09 [-.55; .38]	**.13 [.02; .24]**
SE13→TR14	*-.11[-.22; .00]*	**-.08[-.16; -.01]**	**-.08[-.14; -.02]**	-.02[-.21; .16]	-.12 [-.51; .28]	**.14 [.01; .28]**
SE14→TR15	.01[-.10; .12]	**-.12[-.21; -.02]**	**-.08[-.14; -.03]**	.22[-.01; .44]	-.16 [-.61; .29]	**.14 [-.01; .30]**
SE15→TR16	-.04[-.13; .05]	**-.13[-.23; -.02]**	**-.08[-.13; -.02]**	.04[-.10; .17]	-.20 [-.64; .24]	**.12 [.00; .25]**
TR11→SE12	*.15[-.01; .31]*	**-.07[-.12; -.02]**	-.04[-.09; .01]	.24[-.04; .52]	-.07 [-.46; .33]	.09 [-.18; .37]
TR12→SE13	*-.10[-.20; .01]*	**-.09[-.17; -.02]**	-.05[-.11; .01]	-.05[-.22; .12]	-.08 [-.79; .62]	.10 [-.16; .36]
TR13→SE14	-.07[-.17; .04]	**-.08[-.14; -.01]**	-.06[-.12; .00]	.07[-.15; .30]	-.06 [-.66; .53]	.11 [-.20; .42]
TR14→SE15	-.10[-.22; .01]	**-.06[-.12; -.01]**	-.06[-.13; .01]	.04[-.16; .24]	-.05 [-.51; .41]	.13 [-.22; .47]
TR15→SE16	-.01[-.13; .10]	**-.05[-.10; .00]**	-.06[-.13; .01]	.04[-.10; .19]	-.04 [-.42; .35]	.10 [-.15; .35]

Note. *N*=674; SE = Self-esteem; TR= Teacher rejection. Bold indicates that the 95% confidence interval did not include zero. Italics indicate that the 90% confidence interval did not include zero.

Table S7

Standardized Parameter Estimates [and 95% Confidence Intervals] for Predicting Self-Esteem Four Years Post High School by the Trait Factor of Teacher Support and Rejection Separated by Gender

	Teacher support						Teacher rejection					
	Model 1 Not Controlling. for prior SE		Model 2 Controlling. for SE at Age 11		Model 3 Controlling. for SE at Age 16		Model 1 Not Controlling. for prior SE		Model 2 Controlling. for SE at Age 11		Model 3 Controlling for SE at Age 16	
Predictors	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Trait Factor of TA	**.18 [.06; .31]**	**.26 [.11; .40]**	*.14 [-.01; .28]*	**.21 [.05; .38]**	-.04 [-.13; .12]	.12 [-.03; .28]	**-.15 [-.29; -.02]**	**-.27 [-.42; -.13]**	-.11 [-.26; .04]	**-.22 [-.42; -.01]**	.00 [-.13; .14]	*-.15 [-.30; .01]*
Self-esteem at Age 11			.12 [-.08; .30]	.17 [-.07; .41]					.15 [-.04; .35]	.15 [-.11; .41]		
Self-esteem at Age 16					**.51 [.38; .64]**	**.43 [.30; .56]**					**.51 [.39; .63]**	**.42 [.29; .56]**

Note. *N*=674; SE = Self-esteem; TA= Teacher attachment (support or rejection); Trait factor = Refers to the common variance of teacher support/rejection from age 11 to age 16. Bold indicates that the 95% confidence interval did not include zero. Italics indicate that the 90% confidence interval did not include zero.

Table S8

Standardized Parameter Estimates [and 95% Confidence Intervals] of the (Random Intercept) Cross-Lagged Panel Model (RI-CLPM; CLPM) between Self-Esteem and Teacher Support Separated by Nativity

	CLPM		RI-CLPM	
Path	1st Generation	2nd/3rd Generation	1st Generation	2nd/3rd Generation
Autoregressive				
TS11→TS12	**.47 [.37; .56]**	**.47 [.41; .54]**	**.50 [.36; .64]**	**.25 [.18; .32]**
TS12→TS13	**.50 [.40; .61]**	**.54 [.48; .59]**	**.57 [.45; .68]**	**.29 [.21; .38]**
TS13→TS14	**.55 [.42; .67]**	**.54 [.48; .60]**	**.63 [.50; .76]**	**.30 [.21; .38]**
TS14→TS15	**.50 [.39; .61]**	**.55 [.49; .62]**	**.55 [.41; .69]**	**.32 [.22; .41]**
TS15→TS16	**.55 [.46; .64]**	**.58 [.52; .65]**	**.59 [.47; .70]**	**.34 [.24; .44]**
SE11→SE12	**.66 [.52; .78]**	**.62 [.56; .68]**	.18 [-.04; .40]	**.42 [.31; .54]**
SE12→SE13	**.72 [.61; .84]**	**.65 [.58; .73]**	.19 [-.06; .44]	**.41 [.24; .57]**
SE13→SE14	**.72 [.60; .83]**	**.59 [.53; .66]**	.23 [-.01; .48]	**.37 [.25; .49]**
SE14→SE15	**.72 [.62; .82]**	**.64 [.57; .72]**	.19 [-.08; .46]	**.41 [.24; .57]**
SE15→SE16	**.68 [.57; .78]**	**.61 [.56; .67]**	.19 [-.07; .46]	**.40 [.26; .53]**
Cross-lagged				
SE11→TS12	**.11 [.03; .19]**	**.14 [.08; .20]**	-.10 [-.26; .06]	**.15 [.05; .25]**
SE12→TS13	**.11 [.03; .19]**	**.13 [.08; .19]**	-.10 [-.28; .08]	**.13 [.04; .22]**
SE13→TS14	**.11 [.04; .18]**	**.12 [.07; .17]**	-.11 [-.32; .09]	**.12 [.04; .20]**
SE14→TS15	**.10 [.04; .17]**	**.13 [.08; .17]**	-.09 [-.24; .07]	**.13 [.04; .21]**
SE15→TS16	**.10 [.03; .18]**	**.12 [.08; .17]**	-.09 [-.24; .07]	**.13 [.04; .22]**
TS11→SE12	.04 [-.01; .08]	**.06 [.03; .09]**	.00 [-.16; .16]	*.06 [.00; .11]*
TS12→SE13	.04 [-.01; .09]	**.07 [.03; .11]**	.00 [-.18; .18]	*.07 [.00; .15]*
TS13→SE14	.05 [-.01; .10]	**.07 [.03; .12]**	.00 [-.23; .23]	*.07 [.00; .15]*
TS14→SE15	.04 [-.01; .10]	**.08 [.03; .12]**	.00 [-.21; .21]	*.08 [.00; .17]*
TS15→SE16	.05 [-.01; .10]	**.08 [.03; .12]**	.00 [-.22; .22]	*.08 [.00; .17]*

Note. *N*=674; SE = Self-esteem; TS= Teacher support; 1st generation = born in Mexico; 2nd and 3rd generation = born in the U.S; Models with longitudinal constraints. Bold indicates that the 95% confidence interval did not include zero. Italics indicate that the 90% confidence interval did not include zero.

Table S9

Standardized Parameter Estimates [and 95% Confidence Intervals] of the (Random Intercept) Cross-Lagged Panel Model (RI-CLPM; CLPM) between Self-Esteem and Teacher Rejection Separated by Nativity

	CLPM		RI-CLPM	
Path	1st Generation	2nd/3rdGeneration	1st Generation	2nd/3rd Generation
Autoregressive				
TR11→TR12	.51 [.04; .97]	**.48 [.26; .70]**	.49 [-.52; 1.49]	.02 [-.47; .50]
TR12→TR13	.37 [-.07; .81]	**.64 [.47; .81]**	.37 [-.39; 1.14]	.03 [-.87; .94]
TR13→TR14	.44 [-.07; .94]	**.73 [.55; .91]**	.44 [-.29; 1.16]	.04 [-.91; .98]
TR14→TR15	.33 [-.07; .73]	**.80 [.61; 1.00]**	.25 [-.38; .87]	.03 [-.79; .85]
TR15→TR16	.31 [-.19; .82]	**.69 [.55; .83]**	.32 [-.50; 1.13]	.03 [-.75; .81]
SE11→SE12	**.67 [.53; .80]**	**.61 [.55; .67]**	.22 [-.07; .51]	**.42 [.28; .55]**
SE12→SE13	**.73 [.61; .84]**	**.66 [.58; .73]**	.18 [-.08; .45]	**.37 [.19; .54]**
SE13→SE14	**.73 [.62; .84]**	**.60 [.54; .66]**	.23 [-.04; .49]	**.34 [.21; .47]**
SE14→SE15	**.73 [.63; .83]**	**.65 [.58; .73]**	.18 [-.09; .44]	**.36 [.20; .53]**
SE15→SE16	**.69 [.59; .79]**	**.62 [.56; .68]**	.19 [-.09; .47]	**.35 [.21; .50]**
Cross-Lagged				
SE11→TR12	**-.11 [-.19; -.04]**	-.04 [-.08; .01]	-.01 [-.23; .21]	-.10 [-.29; .10]
SE12→TR13	**-.12 [-.20; -.04]**	-.03 [-.08; .01]	-.01 [-.22; .21]	-.10 [-.30; .11]
SE13→TR14	**-.15 [-.24; -.05]**	-.03 [-.08; .01]	-.01 [-.31; .29]	-.12 [-.40; .16]
SE14→TR15	**-.13 [-.24; -.02]**	-.04 [-.09; .01]	-.01 [-.19; .18]	-.13 [-.42; .15]
SE15→TR16	**-.11 [-.18; -.04]**	-.04 [-.09; .01]	-.01 [-.19; .18]	-.13 [-.38; .13]
TR11→SE12	.03 [-.09; .15]	**-.05 [-.08; -.01]**	-.05 [-.42; .32]	-.03 [-.16; .11]
TR12→SE13	.02 [-.07; .12]	**-.07 [-.13; -.02]**	-.03 [-.28; .21]	-.04 [-.29; .20]
TR13→SE14	.02 [-.07; .11]	**-.07 [-.12; -.03]**	-.04 [-.30; .23]	-.04 [-.23; .16]
TR14→SE15	.02 [-.05; .09]	**-.07 [-.12; -.02]**	-.03 [-.24; .18]	-.03 [-.20; .14]
TR15→SE16	.02 [-.06; .10]	**-.06 [-.10; -.02]**	-.04 [-.33; .25]	-.03 [-.18; .13]

Note. *N*=674; SE = Self-esteem; TS= Teacher rejection; 1st generation = born in Mexico; 2nd and 3rd generation = born in the U.S; Models with longitudinal constraints. Bold indicates that the 95% confidence interval did not include zero.

Table S10

Standardized Parameter Estimates [and 95% Confidence Intervals] for Predicting Self-Esteem Four Years Post High School by the Trait Factor of Teacher Support and Rejection Separated by Nativity

	Teacher support						Teacher rejection					
	Model 1 Not Controlling. For prior SE		Model 2 Controlling. for SE at T1		Model 3 Controlling. for SE at T6		Model 1 Not Controlling for prior SE		Model 2 Controlling. for SE at T1		Model 3 Controlling for SE at T6	
Predictors	1st Gen	2nd/3rd Gen	1st Gen	2nd/3rd Gen	1st Gen	2nd/3rd Gen	1st Gen	2nd/3rd Gen	1st Gen	2nd/3rd Gen	1st Gen	2nd/3rd Gen
Trait Factor of TA	*.14 [-.02; .30]*	**.24 [.12; .35]**	*-.01 [-.26; .24]*	**.20 [.07; .33]**	-.08 [-.25; .10]	*.11 [-.01; .23]*	*-.22 [-.45; .02]*	**-.20 [-.30; -.09]**	-.10 [-.47; .27]	**-.16 [-.28; -.03]**	-.06 [-.31; .18]	-.09 [-.20; .02]
Self-esteem at Age 11			.28 [-.05; .60]	.11 [-.06; .29]					.19 [-.22; .59]	.14 [-.03; .31]		
Self-esteem at Age 16					**.67 [.56; .78]**	**.39 [.27; .51]**					**.64 [.53; .75]**	**.40 [.29; .52]**

Note. *N*=674 SE = Self-esteem; TA= Teacher attachment (support or rejection); Trait factor = Refers to the common variance of teacher support/rejection from age 11 to age 16; 1st generation = born in Mexico; 2nd and 3rd generation = born in the U.S. Bold indicates that the 95% confidence interval did not include zero. Italics indicate that the 90% confidence interval did not include zero.

6 GENERAL DISCUSSION

When studying psychological constructs, researchers need to make important decisions about the methodological implementations of their specific research questions. Improving the fit between theory and methods can produce synergistic effects for scientific progress (Greenwald, 2012; Marsh & Hau, 2007). The present dissertation was aimed at merging theoretical questions about self-esteem with pivotal considerations about the methodological implementations of these questions. Self-esteem is one of the most well-studied constructs in psychology and represents the subjective evaluation of individuals' self-worth (Donnellan et al., 2011). The construct of self-esteem has received a great deal of interest due to its importance for indicators of mental health (Sowislo & Orth, 2013; Trzesniewski et al., 2006). Self-esteem has largely been described as a *unidimensional* and *trait-like* construct, which is predominantly manifested through *social interactions with parents and peers*. However, important deviations and extensions from these well-established and widespread definitions have not been integrated sufficiently, such as the conceptualization of *multidimensional* global self-concept, the consideration *state-like* self-esteem, as well as *other relevant social sources* of students' development such as teachers. Along with theoretical reasons for these gaps, a major challenge involved in these questions is the methodological implementation of these research questions. Therefore, the present dissertation addressed an integration of theory and methods, which led to two overarching objectives. First, this dissertation was aimed at improving the understanding of students' self-esteem during the important developmental periods of adolescence and young adulthood. Specifically, I addressed research questions about the conceptualization and stability of self-esteem as well as individual and environmental relations with other constructs. Second, this dissertation was aimed at improving the understanding of different methodological representations of higher order constructs, states and traits, and reciprocal relations.

These overarching objectives were addressed in three empirical studies. In reference to the first objective, there were three main findings: (a) Global self-concept was best conceptualized as formed by multidimensional lower order self-concepts and was more closely aligned with unidimensional global self-esteem when nonacademic self-concepts were included (Study 1). (b) State and trait self-esteem showed different proportions of state residual and trait variance (i.e., different stability) and this led to different longitudinal relations with depressive symptoms (Study 2). (c) Teachers can indeed be a source of information for students' self-esteem as self-esteem and student-teacher relationships (i.e., teacher attachment) were recipro-

cally related over the course of adolescence and beyond (Study 3). Regarding the second objective, the present dissertation identified the following findings: (a) When lower order constructs were barely correlated, the use of second-order factor models could lead to empirical problems such as low variances or implausible parameter estimates, whereas a model-based latent composite tended to have larger variances and a more consistent pattern of relations (Study 1). (b) Measuring (i.e., using different time frames) and modeling (i.e., by disentangling the components of variance) of states and traits was fundamentally related and had crucial consequences for a bivariate research question (Study 2). (c) Different configurations of cross-lagged models led to different conclusions about the theoretical research questions in the analysis of reciprocal relations (Study 2, Study 3).

The significance, limitations, and implications of these findings will be further discussed in the following sections of Chapter 6. Most importantly, I will apply a broader perspective to address the two overarching goals, thus emphasizing the joint advances of the three empirical studies for theoretical questions in research on self-esteem (Chapter 6.1) and for understanding the role of different methodological implementations (Chapter 6.2). Furthermore, I will discuss limitations and future research (Chapter 6.3), present an outlook for classifying measures (Chapter 6.4), discuss implications for policy and practice (Chapter 6.5), and present an overall conclusion for this dissertation (Chapter 6.6).

6.1 Theoretical Advances for Research on Self-Esteem

The present dissertation investigated three emerging research gaps on the conceptualization, stability, and reciprocal relations of self-esteem. By using variations in the methodological representation of the respective research questions, this dissertation offers new theoretical perspectives on central assumptions about self-esteem. In this overall discussion, I first want to embed findings on the conceptualization of the global self with regard to the *relation between global self-concept and global self-esteem*. Second, I will discuss findings on the stability of self-esteem by covering findings on both *state and trait self-esteem*. Third, the findings of the present dissertation will be classified regarding the knowledge gained for *individual and environmental predictors and consequences.*

6.1.1 The Relation between Global Self-Concept and Global Self-Esteem

Research on the self was grounded in the ideas of William James (1890/1963), who pointed to the "Me" self as representing individuals' self-perceptions, which can refer to different hierarchical levels (i.e., the material, the social, and the spiritual self). This theoretical foundation was split up into different research traditions, which emphasized different aspects of the self. One important differentiation is the unidimensional versus the multidimensional view on the self. Unidimensional views on the self mostly embrace the term global *self-esteem*, whereas multidimensional views on the self most often use the term *self-concept*. The multidimensional perspectives not only conceptualize self-concepts as multidimensional, but they also suggest a hierarchy of multidimensional self-concept with global self-concept at the apex of this hierarchy (Shavelson et al., 1976).

Previous research on multidimensional hierarchical self-concept and self-esteem has been pursued in parallel but has largely been conducted in different disciplines. Whereas fields such as personality and social psychology have most often conducted research on global self-esteem (e.g., Donnellan et al., 2011; Orth & Robins, 2014; Zeigler-Hill, 2013), educational psychology has oftentimes focused on a multidimensional framework of self-concepts, such as self-concepts in different school subjects (e.g., Marsh, 1986, 1987a). Research that has aimed to bring together these disciplines has not been perceived as excessively informative in the individual disciplines: Whereas educational psychologists have tended to downplay the role of self-esteem in education because it could not predict important educational outcomes such as achievement (e.g., Marsh & Craven, 2006; Marsh, Craven et al., 2006; Marsh & O'Mara, 2008; Trautwein, 2003; Trautwein et al., 2006), personality psychologists have not seen the relevance

of including domain-specific self-concepts because domain-specific self-concepts have not been found to contribute sufficiently to global self-esteem (Harris et al., 2018; Orth & Luciano, 2015; Orth & Robins, 2019; Rentzsch et al., 2016) or to predict important life outcomes in the same way as self-esteem (Orth, Robins, Widaman, & Conger, 2014). To some extent, the dissent between multidimensional self-concepts and unidimensional self-esteem could be solved by acknowledging different levels of specificity in the constructs and their target outcomes (Rosenberg, Schooler, Schoenbach, & Rosenberg, 1995; Swann et al., 2007). However, despite the different levels of specificity involved in domain-specific self-concepts and unidimensional global self-esteem, a yet unsolved question is how global self-concept as the most global component of the hierarchy of self-concepts and unidimensional global self-esteem are related. Both global self-concept and global self-esteem represent a conceptualization of the global self, yet the ways in which the global self has been operationalized in these different camps have been somewhat different. Whereas global self-esteem represents a unidimensional construct, which is directly measured (Donnellan et al., 2015; Rosenberg, 1989), global self-concept represents the apex of a self-concept hierarchy, which is based on lower order domain-specific self-concepts (Shavelson et al., 1976). However, in the field of global self-concept, it is still unclear how the operationalization of global self-concept as the apex of the hierarchy should be implemented. On the one hand, a top-down process might be at work, by which global self-concept affects the lower order self-concepts. On the other hand, a bottom-up process might exist, by which the lower order self-concepts form global self-concept.

The theoretical foundations of the hierarchy of self-concept have proposed a bottom-up process by which evaluations in different areas of life form more global perceptions (Shavelson et al., 1976). In stark contrast, empirical studies have mostly applied reflective second-order factor models to represent global self-concept (e.g., Marsh & Shavelson, 1985; Marsh, 1987b), which implies a top-down process. However, the application of second-order factor models was met with empirical problems (i.e., global self-concept had small variances) because the lower order self-concepts were only barely correlated. This led to a neglect of the construct of global self-concept (e.g., Marsh & Shavelson, 1985; Marsh, 1990) rather than the application of different theoretical and methodological approaches for global self-concept (e.g., applying approaches that were more aligned with the theoretically proposed bottom-up formation process). One reason for this paucity might have been that sophisticated methods for implementing a formation process were not available. A recently developed method, namely, the model-based latent composite score (Rose et al., 2019), can now provide this missing piece.

Therefore, the present dissertation laid the foundation for a revival of the construct of global self-concept by representing global self-concept as a model-based latent composite score (Study 1). The present results indicated that this conceptualization of global self-concept offers important advantages over a second-order factor because it resulted in higher variances and more plausible parameter estimates. Moreover, the application of the latent composite score has important implications for evaluating the relation between global self-concept and global self-esteem. Previous research using second-order factors concluded that global self-concept and global self-esteem represented nearly the same construct because they were highly correlated (Marsh & Hattie, 1996). By contrast, the present findings that came about from using a latent composite score to model global self-concept imply a different conclusion. Thus, across the three data sets, global self-concept (based on all available domain-specific self-concepts) and global self-esteem were correlated from only r = .42 to .77. Further, the findings indicated that the relation between global self-concept and global self-esteem was highest when nonacademic self-concepts were included, a finding that is in line with previous findings on the relations between different domain-specific self-concepts and self-esteem (Donnellan, Trzesniewski, Conger, & Conger, 2007; Harter, 2003; Soest et al., 2016). Despite the conjecture that nonacademic self-concepts are more strongly related to global self-esteem than academic self-concepts are, an important subsequent question would be whether academic self-concepts are in any way related to global self-esteem. This question may be complicated by the complexity of the dimensional comparison processes involved in academic self-concepts (e.g., Marsh, 1986).

In sum, the conclusion of the present research is that global self-concept, the apex of the hierarchy of self-concepts, should be conceptualized as a model-based latent composite score. In this regard, an important inference from the present research is that unidimensional global self-esteem is more strongly related to global self-concept when nonacademic self-concepts are included as opposed to academic self-concepts. Thus, self-esteem is not intrinsically equivalent to global self-concept as the apex of the multidimensional hierarchy of self-concept. By contrast, to a large degree, self-esteem seems to represent the nonacademic section of the multidimensional hierarchy of self-concepts, whereas the contribution of academic self-concepts has yet to be clarified.

6.1.2 State and Trait Self-Esteem

Along with different conceptualizations of the global self, previous research has placed particular emphasis on the development of unidimensional global self-esteem. In this regard,

an important differentiation has been made in reference to state and trait self-esteem. According to Leary and Baumeister (2000), state self-esteem refers to short-term thoughts and feelings about the self, whereas trait self-esteem represents enduring thoughts and feelings. Previous longitudinal studies on self-esteem across the lifespan have observed that self-esteem is characterized by large amounts of trait variance (i.e., high rank-order consistency), which is why researchers concluded that self-esteem was a trait-like construct (Orth & Robins, 2014; Trzesniewski et al., 2003). The present dissertation outlined that this conclusion might not only reflect the actual trait-like nature of self-esteem but also depends on how self-esteem is measured. Previous conclusions about the trait-like nature of self-esteem were almost exclusively drawn from measures that were designed to assess trait self-esteem (Rosenberg, 1989). Yet, in line with theoretical assumptions about both trait and state self-esteem, there are also measures that can be used to assess state self-esteem (Heatherton & Polivy, 1991). However, there has been an absence of studies that have simultaneously assessed and compared the consequences of state *and* trait measures of self-esteem.

The present dissertation filled this research gap by experimentally assigning participants to be administered either state or trait measures of self-esteem and subsequently by examining the univariate and bivariate consequences of using these measures in a longitudinal design (Study 2). On the one hand, the findings provided further support for the trait assumption given that the majority of the variance of the state *and* trait measures of self-esteem was explained by a trait factor. On the other hand, the study indicated that, in total, state measures of self-esteem revealed less trait variance than trait measures (and vice versa regarding state residual variance). Thus, these findings experimentally showed the interdependency between state-trait measures and proportions of state-trait variance in self-esteem. Over and above this, the present research indicated that state and trait measures lead to important differences in results on the longitudinal relation between self-esteem and depressive symptoms. Therefore, the findings call for a deeper understanding of the differences in and the relations between state and trait self-esteem. Both state and trait self-esteem can be interesting and relevant sources of self-esteem and have to be set in relation to the theoretical question of interest. In large-scale studies, researchers have typically applied trait measures of self-esteem, which can aid the understanding of the long-term patterns of self-esteem. Yet, when researchers are interested in short-term changes in self-esteem by using more intensive longitudinal data sets, state measures of self-esteem may be more insightful because they are more likely to capture short-term shifts

in self-esteem (e.g., Baumert et al., 2017; Podsakoff, Spoelma, Chawla, & Gabriel, 2019; Vazire & Sherman, 2017; Wrzus & Roberts, 2017).

6.1.3 Relations between Self-Esteem and Individual and Environmental Factors

When studying factors that contribute to the development of psychological constructs, researchers typically draw on both individual and environmental predictors (Bronfenbrenner & Ceci, 1994; Lerner, 1998; Lerner, Lerner, Eye et al., 2011). In research on self-esteem, a variety of individual and environmental predictors have been examined. At the same time, the consequences of different levels of self-esteem can be characterized as either individual or environmental (for an overview of predictors and consequences on self-esteem, see Orth & Robins, 2014, 2019). Interestingly, a broad range of individual and environmental factors have been considered on the sides of both the predictors and the consequences of self-esteem, thus suggesting a reciprocal consideration of these factors when studying their relation to self-esteem. Previous research that has analyzed reciprocal relations has typically relied on traditional cross-lagged panel models, whereas different variations of these models have been proposed to gain deeper insights into relations on the within-person level (Hamaker et al., 2015).

The present dissertation extended the field of individual predictors and consequences by investigating the reciprocal relations between state and trait self-esteem and state and trait depressive symptoms in two samples of university students (Study 2). In line with findings from an extensive field of research investigating the relation between self-esteem and depressive symptoms (for a meta-analysis, see Sowislo & Orth, 2013), the results from Study 2 provide further support for the vulnerability model, which suggests that low self-esteem is a predictor rather than a consequence of depressive symptoms. Over and above previous studies, the present results show that this assumption holds only across trait self-esteem measures and in traditional cross-lagged panel models. By contrast, when using state self-esteem measures and cross-lagged models that control for between-person differences (i.e., a random intercept), the well-researched vulnerability effect did not hold, and the pattern of results even turned in the direction of the scar model to some extent (i.e., depressive symptoms predicted self-esteem). Overall, the present research indicates that the vulnerability model, in which self-esteem is considered a predictor of depressive symptoms, only holds when the typical measure of self-esteem and the typical cross-lagged panel models are used, whereas this pattern did not hold when deviations in measures and models were applied.

In the present dissertation, the topic of environmental predictors and consequences of self-esteem was enhanced by exploring an environmental factor that has rarely been considered previously, namely, students' social relationships with teachers. Due to the strong relevance of social relations in the context of self-esteem (e.g., Cooley, 1902; Leary & Baumeister, 2000), previous research intensively studied the relations between self-esteem and social relationships with parents and peers (Harris & Orth, 2019). Surprisingly, the role of teachers had barely attracted the attention of researchers studying self-esteem development, even though teachers are considered important attachment figures in students' lives (Kesner, 2000). Therefore, the present dissertation targeted this research gap in a 10-year longitudinal study (Study 3). The results consistently showed small reciprocal relations between self-esteem and positive (i.e., teacher support) as well as negative (i.e., teacher rejection) student-teacher relationships in traditional cross-lagged panel models and in similar fashion, but less consistently, reciprocal relations across random intercept cross-lagged panel models. The findings suggest that overall individual differences in student-teacher relationships are reciprocally related to overall individual differences in self-esteem, whereas individual differences in within-person deviations in student-teacher relationships and self-esteem are less consistently related. Theoretically, the findings provide support for the assumption that teachers can be considered significant others who contribute to students' relational value and in turn to their self-esteem, as proposed by sociometer theory (Leary & Baumeister, 2000). At the same time, students' self-esteem prospectively predicts how they perceive interactions with teachers, which would be in line with the self-broadcasting (Srivastava & Beer, 2005) or the risk-regulation perspective (Murray et al., 2006).

Overall, the present dissertation offers important insights into the reciprocal relations between self-esteem and individual as well as environmental factors. The results indicated more consistent results using traditional cross-lagged models compared with models including random intercept factors. Theoretically, this pattern of results could indicate that overall differences rather than within-person deviations account for relations between self-esteem and individual as well as environmental factors. However, in addition, some methodological challenges from models targeting within-person deviations will be further discussion in the next section.

6.2 Advances from and for Integrating Methodological Perspectives

The present dissertation adopted a broad construct validity perspective by comparing *different* methodological representations of substantive research questions on self-esteem. Across the three studies, the results largely indicated that different methodological representations and analytical approaches revealed substantial differences in the results. The methodological variations were driven by theoretical considerations on the conceptualization, stability, and reciprocal relations of self-esteem and therefore, the results were interpreted in reference to these theoretical ideas in the first place. However, from a broader perspective, the results can have important implications for the measurement and modeling of psychological constructs far beyond self-esteem. In this part of the discussion, rather that presenting the results for each of the methodological implementations again, I instead want to delve into the knowledge gained from the joint consideration of the consequences of different methodological implementations across the three studies. Specifically, I want to identify higher order factor models as a common thread across the three studies and discuss the interdependency involved in theory, methods, and data.

6.2.1 A Common Thread: Cross-Sectional and Longitudinal Higher Order Factor Models

Maximizing the fit between theory and methods (e.g., measures or analytical models) needs to be an important prerequisite for drawing conclusions that tie back to the theoretical phenomena of interest (Borsboom et al., 2003; Borsboom, 2006; Marsh & Hau, 2007). The empirical studies from this dissertation provide important insights into the consequences of using different methodological approaches. Overall, in all of these studies, an important focus was on the consequences of higher order factor models. Even if, due to the complexity of the methodological models, this commonality was only implicitly stated in the theoretical background of this dissertation, the common thread of higher order models can now provide important conclusions across the three studies.

On the one hand, a higher order factor was applied cross-sectionally across multidimensional lower order constructs in Study 1 using a *second-order factor*. On the other hand, it was applied longitudinally to multiple latent state variables in Studies 2 and 3 and described as a *trait factor* and/or a *random intercept factor*. Importantly, in all three of the studies, the correlations among the first-order latent variables fundamentally determined the consequences for the research question of interest. In the cross-sectional application in Study 1, low correlations

between the first-order factors led to problematic results in the higher order factor. The longitudinal application in Study 2 showed that state and trait measures influenced the correlations between the latent state variables (i.e., first-order factors) and therefore the amounts of variance in the higher order factor (i.e., trait variance or random intercept). This variance in turn affected the parameter estimates in the different cross-lagged models (Studies 2 and 3). In addition, in Study 3, a longitudinal higher order factor itself (i.e., trait variance or random intercept) was used to predict an outcome variable. In all of these models, the correlation between the first-order factors was decisive because it shaped the parameter estimates and consequently influenced the usefulness of the models when they were applied to address the research questions.

In multidimensional constructs, low correlations between lower order constructs can pose a serious challenge to the usefulness of a higher order factor because they can result in problematic estimates and the invalidity of the construct (Borsboom et al., 2003; Rhemtulla et al., 2019). Therefore, Study 1 provided an alternative theoretical and methodological implementation of multidimensional, hierarchical constructs using a model-based latent composite score, which does not require substantial correlations between the lower order constructs, and, at least in Study 1, it led to more consistent parameter estimates (i.e., variances, stabilities, correlations with external criteria).

For longitudinal studies that model a higher order factor as the trait variance (i.e., random intercept) from multiple latent state variables (Cole, 2012; Steyer et al., 1992), the resulting challenges can be somewhat different. In comparison with multidimensional first-order constructs, there are typically substantial correlations between latent state variables over time, which makes a higher order factor reasonable. However, in some applications of these models, such as in random intercept cross-lagged panel models (Hamaker et al., 2015), researchers focus on "the other end" of a higher order factor model, namely, on the variance that is *not* explained by the higher order factor (i.e., on the latent state residual variance). If a large amount of variance is explained by the higher order factor, the latent state residual variance is comparably small. When using this small state residual variance as a dependent and/or independent variable, such as in random intercept or latent state-trait cross-lagged panel models, this could lead to inaccurate parameter estimates. Study 2 experimentally showed that in the condition in which trait measures were applied, the large amounts of trait variance and subsequently low amounts of state residual variance were not very insightful for understanding a bivariate within-person research question. Hence, researchers who are aiming to control for higher order trait factors (i.e., random intercepts) need to take into consideration the amounts of trait and state

residual variance that are present. This is important not only statistically but also theoretically: If there is not much state residual variance and also not much within-person variability, then models based on within-person deviations (i.e., latent state residual variance) might not be very insightful simply because *there are barely any* within-person deviations. Such constellations of data are likely to appear in large-scale studies that have used global trait-like measures (Baumert et al., 2017). Hence, when researchers are interested in within-person deviations as the entity of interest, they need to consider designing studies that generate substantial within-person data structures using measures that are sensitive enough to detect such fluctuations. In order to do so, it may be particularly promising to assess intensive longitudinal data (Berry & Willoughby, 2017; Orth et al., 2020).

Overall, the findings across the three empirical studies indicated that correlations among first-order factors essentially shaped the consequences of a higher order factor in cross-sectional and longitudinal applications. Whereas low correlations endanger the general usefulness of higher order factors, exceptionally high correlations instead restrict the usefulness of pursuing further research with state residual variances (e.g., within-person deviations).

6.2.2 Integrating Theory, Methods, *and* Data

The present dissertation targeted important theoretical research questions through an integration of theory and methods. The cross-sectional and longitudinal applications of higher order factor models highlighted the consequences for the present research questions given particular constellations of data. Thus, correlations between first-order factors determined the usefulness of complex analytical procedures, such as second-order factor models or different configurations of cross-lagged panel models. By applying *different* methodological perspectives to each of the research questions, the consequences of one specific method could be directly compared with those from another method. In addition, this dissertation also provides insights into the consequences of using different methodological approaches in different data constellations, given that the experimental manipulation of the time frames in Study 2 led to different correlational patterns across the latent state variables. More generally, these findings indicate that the usefulness of particular models depends not only on the theoretical rationale but also on the structure of the data that are collected. In turn, the empirical results from these data and methods can stimulate a rethinking of theory (Gigerenzer, 1991; Greenwald, 2012; Marsh, Byrne, & Yeung, 1999; Orth et al., 2020; Smith, 2005). As such, the present findings point to important theoretical knowledge, for instance, about the formation of global self-concept, the relevance of the vulnerability model in the context of trait but not state self-esteem, and the

limited impact of within-person deviations on the relation between self-esteem and depressive symptoms across 3-week intervals as well as self-esteem and student-teacher relationships in year-to-year intervals.

Along with applying different methodological approaches to gain these theoretical insights, it can be useful to consider scientific progress as a more dynamic process. Even if, in the ideal scenario, researchers choose their analytical models a priori, and their choices are driven by theory, researchers need to expect changes of methods and theory on the basis of empirical data (Marsh et al., 1999). In this regard, even though the preregistration of research questions and analysis plans is an important tool for promoting theory-driven research, to increase transparency, and to reduce publication bias (Nosek, Ebersole, DeHaven, & Mellor, 2018; Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012), at least some analytical "researcher degrees of freedom" (Simmons, Nelson, & Simonsohn, 2011, p. 1359) are often needed in order to adapt to empirical circumstances. Hence, particularly when it comes to complex analytical models, it may be more suitable to think of preregistration as an a priori theory-driven plan of data-driven decisions rather than a fixed a priori decision about which model is appropriate (e.g., Claesen, Gomes, Tuerlinckx, & Vanpaemel, 2019; Nosek et al., 2018; van 't Veer & Giner-Sorolla, 2016). This consideration is important for research fields more generally: Even though it is useful when research fields come to agreements about which theory-driven and sophisticated models are typically used to address certain types of research questions (Orth et al., 2020), a consensus on the usefulness of certain methods should never substitute for research-question-specific considerations about the fit between theory, methods, and data, given that this fit can vary across constructs and/or designs. This problem exposes itself, for example, when journals establish "default" requirements for methods that editors and reviewers ask for. Thus, besides the challenge of optimizing the fit between theoretical questions and their methodological implementation (Borsboom et al., 2003, 2004; Greenwald, 2012; Marsh & Hau, 2007), researchers need to take on the role of a "skilled data detective who follows many alternative leads and makes a case for the most defensible interpretations" (Marsh et al., 1999, p. 163). Hence, the results of the present dissertation should motivate researchers to integrate not only theory and methods but theory, methods, *and* data.

6.3 Limitations and Future Directions

The present dissertation targeted important research questions about self-esteem by bringing them together with different methodological implementations. Across three empirical studies using a total of six independent data sets, this dissertation applied rigorous analytical approaches. This dissertation was written from the perspective of a *quantitatively trained applied researcher* who aimed to take the best of both worlds in order to integrate theory *and* methods. Whereas some might see this as a strength because the integration provided the opportunity to produce advances for the joint understanding of theory and methods, others might not be fully satisfied, for example, because they expected a deeper theoretical perspective or a more fine-grained methodological approach to addressing the research questions. Acknowledging the legitimacy of such concerns, in the following and at a minimum, I want to point out some theoretical and methodological limitations and future directions.

6.3.1 Theoretical Limitations and Future Directions

First, even though the present dissertation provided important insights into the relations between self-concepts and self-esteem, some important aspects of these relations could not be fully clarified. The present dissertation was effective in showing that global self-esteem was more aligned with global self-concept when nonacademic self-concepts were included, whereas academic self-concepts seemed to play a rather minor role. This finding lines up with previous research that indicated that academic self-concepts were barely additively linked to self-esteem (Donnellan et al., 2007; Harris et al., 2018; Rentzsch & Schröder-Abé, 2018). A particularly complex issue that can arise when considering multiple academic self-concepts is the issue of dimensional comparison processes because individuals compare their self-concepts in different domains, and this can result in low or even negative relations between different academic self-concepts (e.g., Marsh, 1986). Therefore, it might be possible that the relations between academic self-concepts and self-esteem involve a more complex mechanism. One explanation could be that academic self-concepts contribute to global self-esteem through a *compensatory* mechanism. For instance, high self-concept in one domain might buffer the effects of low self-concepts in other domains when evaluating the self as a whole. Thereby, it might be possible that low academic self-concepts only impact individuals' self-esteem when there is *not even a single* domain that provides a positive academic experience and mindset. Therefore, future research is needed to investigate more complex patterns of academic self-concepts and their relations to global self-esteem. In order to test a compensatory hypothesis, it might be insightful to analyze interaction effects between particular self-concept domains or person-

centered approaches (e.g., Marsh, Lüdtke, Trautwein et al., 2009) that take into account typical patterns of domain-specific self-concepts in students with high and low self-esteem.

Second, even though the experimental between-subject design used in Study 2 provided important insights into the consequences of state and trait self-esteem, participants answered a measure of either state *or* trait self-esteem, and therefore, the study could not provide knowledge about the relations between state *and* trait self-esteem measures. In recent years, multiple models in personality psychology have provided frameworks for understanding the relations between states and traits (Fleeson & Jayawickreme, 2015; Roberts, 2018; Wrzus & Roberts, 2017). One inference from these models is that it might be promising to connect multiple state measures with single trait measures (Finnigan & Vazire, 2018; Roberts, 2018) in order to understand how short-term state changes in self-esteem contribute to long-term trait self-esteem. For self-esteem, researchers have yet to examine how different state measures contribute to response behavior in trait self-esteem measures. Therefore, an important avenue for future research on self-esteem is to expand the understanding of the theoretical and measurement-specific mechanisms involved in state and trait self-esteem measures, and in this regard, a particular emphasis should lie in understanding how short-term perceptions of self-esteem contribute to more enduring trait self-esteem.

Third, the findings of the three empirical studies enhanced the understanding of the potential influence of individual and environmental factors for the development of self-esteem. On the one hand, this dissertation provided new insights into the reciprocal associations between self-esteem and depressive symptoms. On the other hand, it shed light on the oftentimes neglected social context in school by considering the teacher as an important factor. However, it is important to note that the present studies focused primarily on *one* particular predictor each (i.e., depressive symptoms or student-teacher relationships), and a comprehensive investigation of global self-esteem would require a broader consideration of factors of influence. Therefore, extensions of the present studies applying a broader set of predictors could enrich the understanding of the multideterminacy of self-esteem. One extension for obtaining a broader picture of the factors that are involved in the development of self-esteem would be a more comprehensive consideration of *key developmental contexts.* The school context is a particularly relevant context during youth and adolescence where individuals have diverse experiences, ranging from academic to emotional experiences. It is likely that these different experiences are interwoven when exerting influence on global outcomes of adjustment, such as students' self-esteem. Finally, another extension of the present findings might involve investigating the relation

between individual and/or environmental factors and self-esteem in a *state-trait framework*. Although large-scale studies can provide important insights into the long-term relations between constructs, they are often restricted to year-to-year relations between predictors and outcomes. Studies including more intensive longitudinal data would provide the opportunity to detect associations in smaller pieces. Understanding sets of factors that contribute to state self-esteem can constitute a fruitful starting point for transferring them more broadly to global trait self-esteem.

6.3.2 Methodological Limitations and Future Directions

Even though all of the empirical studies that were conducted as part of this dissertation compared different methodological approaches, by no means does this dissertation offer an all-encompassing methodological representation of the theoretical research questions. Therefore, some methodological considerations need further attention.

First, in order to examine the conceptualizations of higher order constructs, the present dissertation evaluated two fundamental assumptions about the direction between lower and higher order constructs by means of comparing reflective second-order factor models with a newly developed formative approach, namely, the model-based latent composite score. By doing so, this dissertation provided important insights into the consequences of these models when the lower order constructs were barely correlated. Despite this important comparison, each of these methodological implementations clearly represented a simplified depiction of reflective or formative hierarchical constructs. Therefore, it is important to transfer the knowledge gained from the present findings to considerations about more complex higher order constructs and their methodological implementation (e.g., Brunner et al., 2010; Brunner et al., 2012). Moreover, it is important to note that in the present application of the composite score model, we gave equal weights to all components, whereas in the second-order factor model, we estimated the weights (i.e., factor loadings). Despite the fact that equal weights can be a plausible assumption (Wainer, 1976), future studies drawing on a composite score could further improve the theoretical relevance of the composite score by theoretically deriving and assigning weights to the components.

Second, the results from the present dissertation emphasized the importance of using state and trait time frames to assess the constructs. The time frames were consequential not only for decomposing the state-trait variance in the constructs but also for determining the

direction of effects in the bivariate research question. As the first study to experimentally manipulate time frames in measures and test them to address a longitudinal research question, this research represents an important foundation for potential effects involved in the variation of time frames. However, the present study was limited to two constructs and two time frames, which can provide only initial access to understanding the consequences of time frames. Yet, this research needs to be extended by studying additional time frames (e.g., "In the moment," "Today," "During the past year") in a variety of different constructs. In this regard, a more systematic classification of different time frames could stimulate in-depth considerations about the characteristics and consequences of particular measures.

Third, the present dissertation investigated reciprocal relations between constructs using three different types of cross-lagged models. By doing so, the findings substantially contributed to the understanding of the consequences involved in modeling a random intercept to analyze cross-lagged effects. However, other models have been developed to investigate reciprocal relations. These models can additionally model, for example, between-person differences in the trajectories over time (e.g., Bollen & Curran, 2004; Curran, Howard, Bainter, Lane, & McGinley, 2014). An application of these models to the present research questions and data sets (e.g., the different time frame conditions) could expand the knowledge gained in the present dissertation. Besides applying other types of cross-lagged models, it is important to mention that some of the theoretical conclusions drawn from the empirical studies using different configurations of cross-lagged models *implicitly* pointed to causal relations (Grosz, Rohrer, & Thoemmes, in press). Whereas this may be adequate for comparing cross-lagged results across the experimental manipulation of time frames, it may be controversial for the theoretical interpretation of prospective relations between constructs over time because the studies were not experimental, nor did they adequately consider time-invariant or time-varying third variables (Usami, Murayama, et al., 2019). Thus, even though studying these research questions may stimulate researchers to extend the understanding of prospective relations over time, those conducting future research may wish to adopt analytical approaches that consider time-invariant as well as time-varying covariates, such as marginal structural models (Robins, Hernán, & Brumback, 2000).

6.4 An Outlook: A Time Frame – Specificity Classification

The present dissertation assessed students' multidimensional self-concepts and unidimensional global self-esteem and differentiated between trait and state measures of self-esteem. Broadly speaking, all of these measures targeted individuals' self-evaluations. However, these findings indicate that important unique characteristics result when different approaches are applied to measure such constructs. Thereby, they emphasize the need for a more systematic consideration of general characteristics of measures. Stimulated by the present investigations on self-concepts and self-esteem, in this outlook, I thus want to adopt a more systematic perspective on two characteristics of measures, which can have important implications over and above self-concepts and self-esteem.

In general, when assessing psychological constructs such as the characteristics, feelings, and (typical) behavior of individuals, psychological researchers often apply self-report measures in which individuals have to rate a number of statements. Apart from self-esteem and self-concept, this applies to a broad range of socioemotional constructs such as other personality constructs (e.g., the Big Five, life satisfaction), motivation (e.g., value beliefs, interests), or clinical symptoms (e.g., depressive symptoms, anxiety). Using self-esteem and self-concepts as examples, the findings and considerations of the present dissertation enhanced the understanding of two characteristics of items, which are interwoven with the definition of the theoretical construct: on the one hand, the items' *specificity*, which defines the level of abstraction in the theoretical construct, and on the other hand, the items' *time frame*, which represents the definition of the consistency of the theoretical construct. Accordingly, most self-report items can be placed somewhere on a Time Frame – Specificity Classification matrix (see Figure 4). The more global an item is in terms of specificity and/or time frame, the more knowledge participants must integrate to answer the item. Every item is implicitly or explicitly embedded in a certain time frame and entails a certain level of specificity. For example, the Rosenberg Self-Esteem Scale (Rosenberg, 1989) includes the most global time frame and the most global level of specificity as it refers to unspecific, global feelings about the self (e.g., "All in all, I am satisfied with myself"). In the following, I explain the characteristics of these two continua in more detail, which is followed by several suggestions for how such a classification can be applied to improve future research.

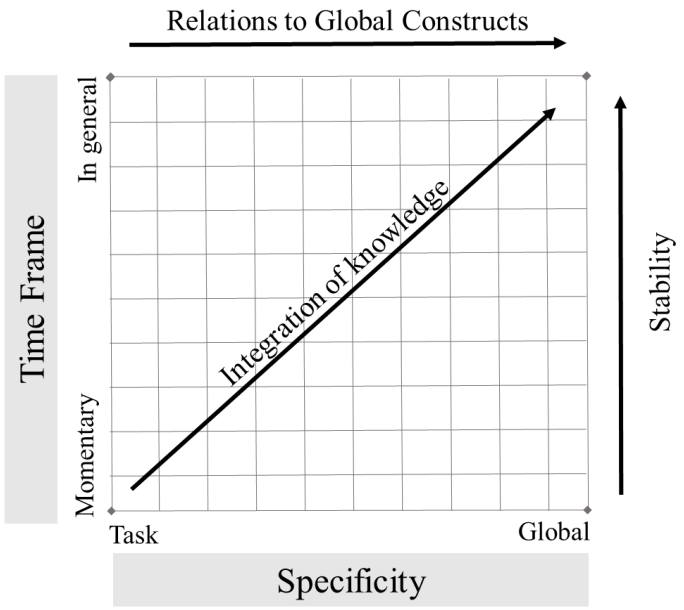


Figure 4. Time Frame – Specificity Classification

Time Frame Continuum. Time frames of items typically range from "momentary" to "in general," where "momentary" presumably represents the most state-like time frame and "in general" the most trait-like time frame. Whereas there is agreement that "in general" refers to a trait-like measure (Robinson & Clore, 2002), the definition of state-like measures is not as clear-cut. In the wake of a specific construct, researchers need to decide which time frame is informative for their specific construct and research question of interest. Previous research on positive and negative affect and anxiety (Lance et al., 2019; Watson et al., 1988) as well as the findings from Study 2 on self-esteem and depressive symptoms indicated that the broader the time frame is, the more stable the construct is over time. These findings provide a rough outline of the relation between time frames and stability. However, a more granulated and systematic analysis using a broader range of constructs and time frames is needed. When an item on a questionnaire does not identify a specific time frame, at best, it can be assumed that individuals refer to the most global time frame of "In general…." However, at worst, it is possible that each and every person will answer in reference to a unique time frame when one is not specified.

Specificity Continuum. The specificity level of items typically ranges from very specific, such as regarding a specific task (or a subdomain, depending on the construct) to very global. Assessments of many socioemotional constructs exist at more than only one point on this continuum, and, to a certain degree, may vary across different parts of the continuum. The

self has been a prominent example that researchers have pointed to and for which they have assessed different levels of specificity (Bandura, 1999; Bong & Skaalvik, 2003; Jansen & Lösch, in prep, as cited in Lösch, 2016; Shavelson et al., 1976). Beyond the self, considerations about the level of specificity have been made in reference to many constructs, for example, in educational and personality psychology. For instance, the Big Five personality traits (McCrae & Costa, 2008; Mõttus, Kandler, Bleidorn, Riemann, & McCrae, 2017; Mõttus et al., 2019) or motivational constructs such as value beliefs (e.g., Dietrich, Viljaranta, Moeller, & Kracke, 2017; Gaspard, 2015) or interests (Hidi & Renninger, 2006) have been described and assessed at different levels of specificity. A classification of measures with respect to the specificity level therefore seems applicable to a broad range of constructs. According to the specificity matching principle (Swann et al., 2007; see also Brunswik symmetry, Brunswik, 1955; Wittmann, 1988), relations between different constructs are stronger when constructs are arranged around the same area of specificity on this matrix. Therefore, the more global the constructs are, the more strongly they should be related to comparably global constructs.

The Time Frame – Specificity classification matrix should represent a descriptive framework for measures and should be able to provide several advantages for future research:

(a) This classification could **foster considerations** about the time frame and specificity of items, which should be fundamentally related to the theoretical definition of the construct.

(b) This classification could urge researchers to make **more precise predictions about the stability** of constructs given that items with more global time frames are typically more stable over time. On the specificity continuum, so far, no clear predictions about the stability of constructs can be made. Even though, for hierarchical self-concept, Shavelson et al. (1976) proposed that more global self-concepts were more stable over time (compared with more specific constructs), empirical findings did not corroborate this assumption (e.g., Marsh, 1987b; Marsh & Hattie, 1996; Rentzsch & Schröder-Abé, 2018). More research is needed to investigate whether an interdependency between specificity and stability can be found for other constructs. Moreover, systematic research is needed to investigate the consequences of combinations of specificity and time frame for the stability of constructs.

(c) Based on this classification, **more precise assumptions about the relations between predictors and consequences** could be made given the specificity of the constructs (Swann et al., 2007). A similar consideration may also be important for time frames. Even though Study 2 did not find a pattern that clearly indicated that constructs assessed with the

same time frame were more strongly related than constructs with different time frames, the findings from Study 2 still showed that the bivariate relations differed as a function of the time frames of the measures. Along with the unique contributions of the time frame and the specificity level, more research is needed to investigate the effects of particular combinations of time frames and specificity.

(d) This classification could encourage researchers to **systematically study variations** on both the specificity and the time frame continuum as well as interactions of the continua. Study 2 showed that experimental variations in time frames were fundamentally related to univariate and bivariate results. This study could be a starting point for more (experimental) research studying univariate and bivariate patterns of results across multiple levels of time frames and specificity in a broad range of constructs. In this regard, an important extension would be to zoom in on differences in the response processes involved in different time frames and different specificity levels. Survey research has suggested that the amount of knowledge individuals potentially need to integrate when answering an item can affect the accuracy and completeness of the response behavior (Jobe, Tourangeau, & Smith, 1993) as well the use of semantic rather than episodic knowledge (Robinson & Clore, 2002). These processes require further attention because they can have important consequences for the theoretical constructs.

In sum, a classification system for items regarding the time frame and specificity brings together many considerations about the nature of psychological constructs raised in the present dissertation and could encourage future research to more deeply consider the characteristics and consequences of the time frame and specificity of their measures.

6.5 Implications for the Construct of Global Self-Esteem in Policy and Practice

Because of its relevance for individuals' mental health, the construct of self-esteem has received a great deal of attention in research as well as in policy and practice. In the 1980s, the state of California even founded a task force to promote self-esteem. In this regard, self-esteem was considered a "social vaccine" (California State Department of Education, 1990, p. 10), which was expected to protect people from a broad range of negative outcomes including not only mental problems but also academic failure, family problems, teenage pregnancy, drug and alcohol use, and poverty. Even though the government of California was strongly criticized for their rather ill-defined, undifferentiated, and ideologized policy (e.g., Baumeister et al., 2003; Kahne, 1996; Swann, 1996), their ultimate goal to improve a society's social and emotional makeup, including the development of a healthy sense of self-worth, still holds. For example, after many years of performance orientation as the primary goal for industrialized countries, more recent claims brought back to the table the notion that improving students' socioemotional skills should be a key policy issue for countries (e.g., Bleidorn et al., 2019; Chernyshenko, Kankaraš, & Drasgow, OECD, 2017a, 2017b).

The importance of improving self-esteem is particularly grounded in the risks associated with low self-esteem. Most prominently, low self-esteem has been described as a vulnerability factor for psychological adjustment, for instance, for depression (Orth et al., 2008; Sowislo & Orth, 2013). The present research further supported this assumption in a sample of freshmen students right after they transitioned to a university. Here, students with low longstanding self-esteem were at particularly high risk of exhibiting depressive symptoms. These findings result in the question of how to prevent students from developing low self-esteem. It is likely that this is a particularly demanding undertaking because global self-esteem is one of the broadest self-evaluations. Meta-analyses conducted across a broad range of different intervention programs and designs have indicated that there were small effects of interventions on global self-esteem, whereas the effects were particularly pronounced for individuals and samples with self-esteem deficits (Haney & Durlak, 1998; O'Mara, Marsh, Craven, & Debus, 2006). In this dissertation, three empirical studies targeted emerging research questions about global self-esteem. Even though it should be noted that this dissertation was primarily aimed at understanding *fundamental* psychological processes, which do not directly imply recommendations for practical actions, some of the findings can provide an important starting point for reducing the risk of low self-esteem in students. The phases of adolescence and young adulthood seem to be particularly important periods for individuals' self-esteem because in

these phases, self-evaluations can go through stormy times due to developmental changes and new individual and societal challenges (Arnett, 1999; Rosenberg, 1986). The present dissertation suggests that during this phase, it would be beneficial to approach global self-esteem with a bottom-up approach. Thus, rather than directly aiming to change global self-evaluation, it might be more expedient to address components and factors of influence that contribute to the *formation* of global self-esteem (see also Harter, 1999). In particular, nonacademic perceptions seem to contribute to feelings about the global self, such as self-perceptions of social relations. Along with social support from parents and peers, the present dissertation emphasized the role of teachers in the development of self-esteem. Thus, teachers who hold positive and trustworthy relationships with their students might contribute to students' general feelings about the self. Along with positive contributions of nonacademic experiences in school, by contrast, the role of academic experiences in the formation of students' global self-esteem seems to be a more complex processes because, typically, individuals' perceptions of their academic experiences in different school subjects underlie dimensional comparison processes. The attempt to enhance self-esteem by improving academic self-concept could therefore lead to conflicting effects in that increasing self-perceptions in one academic domain might lower perceptions in another domain (Möller & Köller, 2001).

Overall, this dissertation corroborated findings from previous studies on risks of low self-esteem in the emergence of depressive symptoms. Moreover, the present findings suggest that rather than directly addressing the global self, researchers, practitioners, and policymakers should approach the malleability of students' self-esteem from a formation perspective by targeting factors that form self-esteem. This means that it may prove to be particularly useful to focus on the antecedents of self-esteem. In this regard, the present results point to the relevance of students' nonacademic experiences, for instance student-teacher relationships. However, prior to an overly hasty translation of these findings into intervention programs and policies, more applied research to explore the conditions and processes involved in the formation of self-esteem is needed.

6.6 Conclusion

The present dissertation extended the understanding of the conceptualization, stability, and reciprocal relations of global self-esteem by integrating theory and methods. It has thereby generated important advances about the formation of global self-concept, state and trait self-esteem, and the relations between self-esteem and individual (i.e., depressive symptoms) and environmental (i.e., student-teacher relationships) factors. Over and above research on self-esteem, the present considerations and empirical findings should convince researchers that it is not only worthwhile but also necessary to occasionally leave the "comfort zone" of well-established assumptions and state-of-the-art methods. In order to understand human affect, behavior, and cognition, it is an ongoing challenge to constantly evaluate the fit between theories, methods, and data. Even for constructs such as self-esteem—one of the oldest and most well-studied constructs in psychology—this challenge is ever-present.

7 REFERENCES

Ainsworth, M. D.S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment*. Hillsdale, NJ: Erlbaum Press. Retrieved from http://gbv.eblib.com/patron/FullRecord.aspx?p=1596620

Aldrup, K., Klusmann, U., Lüdtke, O., Göllner, R., & Trautwein, U. (2018). Social support and classroom management are related to secondary students' general school adjustment: A multilevel structural equation model using student and teacher ratings. *Journal of Educational Psychology*, *110*(8), 1066–1083. https://doi.org/10.1037/edu0000256

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.).* Washingtion, DC: Author.

Arnett, J. J. (1999). Adolescent storm and stress, reconsidered. *American Psychologist*, *54*(5), 317–326. https://doi.org/10.1037//0003-066x.54.5.317

Baldwin, J. M. (1895). *Mental development of the child and the race: Methods and process*. New York: Macmillan.

Bandura, A. (1999). A social cognitive theory of personality. In L. Pelvin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 154–196). New York: Guilford Press.

Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest*, *4*(1), 1–44. https://doi.org/10.1111/1529-1006.01431

Baumert, A., Schmitt, M., Perugini, M., Johnson, W., Blum, G., Borkenau, P.,. . . Mõttus, R. (2017). Integrating personality structure, personality process, and personality development. *European Journal of Personality*, *31*(5), 503–528. https://doi.org/10.1002/per.2115

Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. New York, NY: Harper & Row.

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

Biesanz, J. C. (2012). Autoregressive longitudinal models. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 459–471). New York, NY: Guilford Press Pubn.

Blalock, H. M. (1964). *Causal inferences in nonexperimental research*. Chapel Hill: University of North Carolina Press.

Bleidorn, W., Arslan, R. C., Denissen, J. J. A., Rentfrow, P. J., Gebauer, J. E., Potter, J., & Gosling, S. D. (2016). Age and gender differences in self-esteem-A cross-cultural window. *Journal of Personality and Social Psychology*, *111*(3), 396–410. https://doi.org/10.1037/pspp0000078

Bleidorn, W., Hill, P. L., Back, M. D., Denissen, J. J. A., Hennecke, M., Hopwood, C. J.,. . . Roberts, B. (2019). The policy relevance of personality traits. *American Psychologist*, *74*(9), 1056–1067. https://doi.org/10.1037/amp0000503

Bleidorn, W., Hufer, A., Kandler, C., Hopwood, C. J., & Riemann, R. (2018). A nuclear twin family study of self-esteem. *European Journal of Personality*, *32*(3), 221–232. https://doi.org/10.1002/per.2136

Bleidorn, W., Schönbrodt, F., Gebauer, J. E., Rentfrow, P. J., Potter, J., & Gosling, S. D. (2016). To live among like-minded others: Exploring the links between person-city personality fit and self-esteem. *Psychological Science*, *27*(3), 419–427. https://doi.org/10.1177/0956797615627133

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*(2), 305–314. https://doi.org/10.1037/0033-2909.110.2.305

Bollen, K. A., & Bauldry, S. (2011). Three Cs in measurement models: causal indicators, composite indicators, and covariates. *Psychological Methods*, *16*(3), 265–284. https://doi.org/10.1037/a0024448

Bollen, K. A., & Curran, P. J. (2004). Autoregressive Latent Trajectory (ALT) models: A synthesis of two traditions. *Sociological Methods & Research*, *32*(3), 336–383. https://doi.org/10.1177/0049124103260222

Bollen, K. A., & Davis, W. R. (2009). Causal indicator models: Identification, estimation, and testing. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 498–522. https://doi.org/10.1080/10705510903008253

Bollen, K. A., & Diamantopoulos, A. (2017). In defense of causal-formative indicators: A minority report. *Psychological Methods*, *22*(3), 581–596. https://doi.org/10.1037/met0000056

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

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika, 71*(3), 425–440. https://doi.org/10.1007/s11336-006-1447-6

Borsboom, D., Cramer, A., Kievit, R., Zand Scholten, A., & Franic, S. (2009). The end of construct validity. In R. W. Lissitz (Ed.), *The concept of validity. Revisions, New Directions and Applications* (pp. 135–170). Charlotte, NC: Information Age Publishing.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*(2), 203–219. https://doi.org/10.1037/0033-295X.110.2.203

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Bowlby, J. (1969). *Attachment and loss: Vol. 1. Attachment*. New York, NY: Basic Books.

Bowlby, J. (1973). *Attachment and loss: Vol. 2. Separation: Anxiety and anger*. New York, NY: Basic Books.

Bowlby, J. (1980). *Attachment and loss: Vol. 3. Loss: Sadness and depression*. New York, NY: Basic Books.

Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nurture reconceptualized in developmental perspective: a bioecological model. *Psychological Review*, *101*(4), 568–586. https://doi.org/10.1037/0033-295x.101.4.568

Brunner, M., Keller, U., Dierendonck, C., Reichert, M., Ugen, S., Fischbach, A., & Martin, R. (2010). The structure of academic self-concepts revisited: The nested Marsh/Shavelson model. *Journal of Educational Psychology*, *102*(4), 964–981. https://doi.org/10.1037/a0019644

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, *80*(4), 796–846. https://doi.org/10.1111/j.1467-6494.2011.00749.x

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193–217. https://doi.org/10.1037/h0047470

California State Department of Education. (1990). *Toward a state of self-esteem: The final report of the California Task Force to Promote Self-Esteem and Personal Responsibility*. Sacramento, CA: Author. Retrieved from https://files.eric.ed.gov/fulltext/ED321170.pdf

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multi-trait-multimethod matrix. *Psychological Bulletin, 56*(2), 81–105. https://doi.org/10.1037/h0046016

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: stability and change. *Annual Review of Psychology*, *56*, 453–484. https://doi.org/10.1146/annurev.psych.55.090902.141913

Cattell, R. B. (1940). A culture-free intelligence test, I. *Journal of Educational Psychology*. (31), 161–179.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*. (54), 1–22.

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A Comparison of Bifactor and Second-Order Models of Quality of Life. *Multivariate Behavioral Research, 41*(2), 189–225. https://doi.org/10.1207/s15327906mbr4102_5

Chernyshenko, O., Kankaraš, M., & Drasgow, F. *Social and emotional skills for student success and well-being: Conceptual framework for the OECD study on social and emotional skills. OECD Education Working Papers* (Vol. 173). Paris: OECD Publishing.

Claesen, A., Gomes, S. L. B. T., Tuerlinckx, F., & Vanpaemel, W. (2019). Preregistration: Comparing dream to reality. Retrieved from https://psyarxiv.com/d8wex/

Cole. (2012). Latent state-trait models. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling.* New York, NY: Guilford Press Pubn.

Cole, D. A., Martin, N. C., & Steiger, J. H. (2005). Empirical and conceptual problems with longitudinal trait-state models: introducing a trait-state-occasion model. *Psychological Methods, 10*(1), 3–20. https://doi.org/10.1037/1082-989X.10.1.3

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.

Cooley, C. H. (1902). *Human nature and the social order*. New York: Charles Scribner´s Sons.

Coopersmith, S. (1967). *The antecedents of self-esteem*. San Francisco: W. H. Freeman.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

Curran, P. J., Howard, A. L., Bainter, S. A., Lane, S. T., & McGinley, J. S. (2014). The separation of between-person and within-person components of individual change over time: a latent curve model with structured residuals. *Journal of Consulting and Clinical Psychology*, *82*(5), 879–894. https://doi.org/10.1037/a0035297

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Springer Science+Business Media.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, *61*(12), 1203–1218. https://doi.org/10.1016/j.jbusres.2008.01.009

Dietrich, J., Viljaranta, J., Moeller, J., & Kracke, B. (2017). Situational expectancies and task values: Associations with students' effort. *Learning and Instruction*, *47*, 53–64. https://doi.org/10.1016/j.learninstruc.2016.10.009

Donnellan, M. B., Kenny, D. A., Trzesniewski, K. H., Lucas, R. E., & Conger, R. D. (2012). Using trait-state models to evaluate the longitudinal consistency of global self-esteem from adolescence to adulthood. *Journal of Research in Personality*, *46*(6), 634–645. https://doi.org/10.1016/j.jrp.2012.07.005

Donnellan, M. B., Trzesniewski, K. H., Conger, K. J., & Conger, R. D. (2007). A three-wave longitudinal study of self-evaluations during young adulthood. *Journal of Research in Personality*, *41*(2), 453–472. https://doi.org/10.1016/j.jrp.2006.06.004

Donnellan, M. B., Trzesniewski, K. H., & Robins, R. W. (2011). Self-esteem: Enduring issues and controversies. In T. Chamorro-Premuzic (Ed.), *Wiley-Blackwell handbooks in personality and individual differences: Vol. 1. The Wiley-Blackwell handbook of individual differences* (pp. 718–746). Chichester: Wiley-Blackwell.

Donnellan, M. B., Trzesniewski, K. H., & Robins, R. W. (2015). Measures of self-esteem. In G. J. Boyle, D. H. Saklofske, & G. Matthews (Eds.), *Measures of personality and social psychological constructs* (pp. 131–157). London [etc.]: Academic Press. https://doi.org/10.1016/B978-0-12-386915-9.00006-1

Donnellan, M. B., Trzesniewski, K. H., Robins, R. W., Moffitt, T. E., & Caspi, A. (2005). Low self-esteem is related to aggression, antisocial behavior, and delinquency. *Psychological Science, 16*(4), 328–335. https://doi.org/10.1111/j.0956-7976.2005.01535.x

Eccles, J., & Midgley, C. (1989). Stage-environment fit: Developmentally appropriate classrooms for young adolescents. In Ames, C., Ames, R. (Ed.), *Research on motivation in education: Vol. 3. Goals and cognitions* (pp. 139–186). New York, NY: Academic Press.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155–174. https://doi.org/10.1037/1082-989X.5.2.155

Eid, M., Holtmann, J., Santangelo, P., & Ebner-Priemer, U. (2017). On the definition of latent-state-trait models with autoregressive effects. *European Journal of Psychological Assessment*, *33*(4), 285–295. https://doi.org/10.1027/1015-5759/a000435

Epstein, S. (1973). The self-concept revisited. Or a theory of a theory. *American Psychologist*, *28*(5), 404–416. https://doi.org/10.1037/h0034679

Epstein, S., & O'Brien, E. J. (1985). The person–situation debate in historical and current perspective. *Psychological Bulletin*, *98*(3), 513–537. https://doi.org/10.1037/0033-2909.98.3.513

Erol, R. Y., & Orth, U. (2011). Self-esteem development from age 14 to 30 years: A longitudinal study. *Journal of Personality and Social Psychology*, *101*(3), 607–619. https://doi.org/10.1037/a0024299

Eysenck, H. J. (1983). Cicero and the state-trait theory of anxiety: Another case of delayed recognition. *American Psychologist*, *38*(1), 114–115. https://doi.org/10.1037/0003-066X.38.1.114

Finnigan, K. M., & Vazire, S. (2018). The incremental validity of average state self-reports over global self-reports of personality. *Journal of Personality and Social Psychology*, *115*(2), 321–337. https://doi.org/10.1037/pspp0000136

Fitts, W. H. (1965). *A manual for the Tennessee Self-Concept Scale*. Nashville, TN: Counselor Recordings and Tests.

Fleeson, W., & Jayawickreme, E. (2015). Whole trait theory. *Journal of Research in Personality, 56*, 82–92. https://doi.org/10.1016/j.jrp.2014.10.009

Gallagher, M. W., Lopez, S. J., & Preacher, K. J. (2009). The hierarchical structure of well-being. *Journal of Personality*, *77*(4), 1025–1050. https://doi.org/10.1111/j.1467-6494.2009.00573.x

Gaspard, H. (2015). *Promoting value beliefs in mathematics: A multidimensional perspective and the role of gender (Doctoral dissertation, University of Tübingen)*. Retrieved from http://hdl.handle.net/10900/63819

Geiser, C., Götz, T., Preckel, F., & Freund, P. A. (2017). States and traits. *European Journal of Psychological Assessment*, *33*(4), 219–223. https://doi.org/10.1027/1015-5759/a000413

Geukes, K., Nestler, S., Hutteman, R., Dufner, M., Küfner, A. C. P., Egloff, B.,. . . Back, M. D. (2017). Puffed-up but shaky selves: State self-esteem level and variability in narcissists. *Journal of Personality and Social Psychology*, *112*(5), 769–786. https://doi.org/10.1037/pspp0000093

Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, *98*(2), 254–267. https://doi.org/10.1037/0033-295X.98.2.254

Greenwald, A. G. (2012). There is nothing so theoretical as a good method. *Perspectives on Psychological Science*, *7*(2), 99–108. https://doi.org/10.1177/1745691611434210

Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (in press). The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science.* Retrieved from https://doi.org/10.31234/osf.io/8hr7n

Gruenenfelder-Steiger, A. E., Harris, M. A., & Fend, H. A. (2016). Subjective and objective peer approval evaluations and self-esteem development: A test of reciprocal, prospective, and long-term effects. *Developmental Psychology*, *52*(10), 1563–1577. https://doi.org/10.1037/dev0000147

Gustafsson, J. E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, *28*(4), 407–434. https://doi.org/10.1207/s15327906mbr2804_2

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, *8*(3), 179–203. https://doi.org/10.1016/0160-2896(84)90008-4

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

Haney, P., & Durlak, J. A. (1998). Changing self-esteem in children and adolescents: a meta-analytic review. *Journal of Clinical Child Psychology*, *27*(4), 423–433. https://doi.org/10.1207/s15374424jccp2704_6

Harris, M. A., & Orth, U. (2019). The link between self-esteem and social relationships: A meta-analysis of longitudinal studies. *Journal of Personality and Social Psychology*. Advance online publication. https://doi.org/10.1037/pspp0000265

Harris, M. A., Wetzel, E., Robins, R. W., Donnellan, M. B., & Trzesniewski, K. H. (2018). The development of global and domain self-esteem from ages 10 to 16 for Mexican-origin youth. *International Journal of Behavioral Development*, *42*(1), 4–16. https://doi.org/10.1177/0165025416679744

Harter, S. (1983). Developmental perspectives on the self-system. In E. M. Hetherington (Ed.), *Handbook of child psychology. Socialization, personality and social development* (4th ed., pp. 275–386). New York: Wiley.

Harter, S. (1990). Processes underlying adolescent self-concept formation. In R. Montemayor, G. Adams, & T. Gullotta (Eds.), *From childhood to adolescence: A transitional period?* (pp. 205–239). Thousand Oaks, CA: Sage.

Harter, S. (1993). Causes and conequences of low self-esteem in children and adolescents. In R. F. Baumeister (Ed.), *Self-esteem: The puzzle of low self-regard* (pp. 87–116). New York, NY: Plenum Press.

Harter, S. (1998). The development of self-representations. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Vol 3. Social, emotional, and personality development* (5th ed., pp. 553–617). New York: Wiley.

Harter, S. (2003). The development of self-representations during childhood and adolescence. In In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 610–642). New York: Wiley.

Harter, S. (1999). *The construction of the self: A developmental perspective*. *Distinguished contributions in psychology*. New York: Guilford Press. Retrieved from http://www.loc.gov/catdir/bios/guilford051/98056206.html

Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology*, *60*(6), 895–910. https://doi.org/10.1037/0022-3514.60.6.895

Hertzog, C., & Nesselroade, J. R. (1987). Beyond autoregressive models: Some implications of the trait-state distinction for the structural modeling of developmental change. *Child Development*, *58*(1), 93. https://doi.org/10.2307/1130294

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Hutteman, R., Nestler, S., Wagner, J., Egloff, B., & Back, M. D. (2015). Wherever I may roam: Processes of self-esteem development from adolescence to emerging adulthood in the context of international student exchange. *Journal of Personality and Social Psychology*, *108*(5), 767–783. https://doi.org/10.1037/pspp0000015

James, W. (1890/1963). *The principles of psychology*. New York: Holt, Rinehart & Winston.

Jansen, M., & Lösch, T. (in prep). *Integrating perspectives on the self-appraisal of academic abilities: A classification model [Manuscript in preperation]*.

Jobe, J. B., Tourangeau, R., & Smith, A. F. (1993). Contributions of survey research to the understanding of memory. *Applied Cognitive Psychology*, *7*(7), 567–584. https://doi.org/10.1002/acp.2350070703

Kahne, J. (1996). The politics of self-esteem. *American Educational Research Journal*, *33*(1), 3–22. https://doi.org/10.3102/00028312033001003

Kenny, D. A., & Zautra, A. (1995). The trait-state-error model for multiwave data. *Journal of Consulting and Clinical Psychology*, *63*(1), 52–59. https://doi.org/10.1037/0022-006X.63.1.52

Kenny, D. A., & Zautra, A. (2001). Trait–state models for longitudinal data. In L. M. Collins (Ed.), *Decade of behavior. New methods for the analysis of change* (pp. 243–263). Washington D.C.: American Psychological Association. https://doi.org/10.1037/10409-008

Kernis, M. H. (2005). Measuring self-esteem in context: The importance of stability of self-esteem in psychological functioning. *Journal of Personality*, *73*(6), 1569–1605. https://doi.org/10.1111/j.1467-6494.2005.00359.x

Kernis, M. H., Cornell, D. P., Sun, C.-R., Berry, A., & Harlow, T. (1993). There's more to self-esteem than whether it is high or low: The importance of stability of self-esteem. *Journal of Personality and Social Psychology*, *65*(6), 1190–1204. https://doi.org/10.1037/0022-3514.65.6.1190

Kesner, J. E. (2000). Teacher characteristics and the quality of child–teacher relationships. *Journal of School Psychology*, *38*(2), 133–149. https://doi.org/10.1016/S0022-4405(99)00043-6

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 593–602. https://doi.org/10.1001/archpsyc.62.6.593

Kieling, C., Adewuya, A., Fisher, H. L., Karmacharya, R., Kohrt, B. A., Swartz, J. R., & Mondelli, V. (2019). Identifying depression early in adolescence. *The Lancet Child & Adolescent Health*, *3*(4), 211–213. https://doi.org/10.1016/S2352-4642(19)30059-8

Kling, K. C., Hyde, J. S., Showers, C. J., & Buswell, B. N. (1999). Gender differences in self-esteem: A meta-analysis. *Psychological Bulletin*, *125*(4), 470–500. https://doi.org/10.1037/0033-2909.125.4.470

Krauss, S., Orth, U., & Robins, R. W. (2019). Family environment and self-esteem development: A longitudinal study from age 10 to 16. *Journal of Personality and Social Psychology.* Advance online publication. https://doi.org/10.1037/pspp0000263

Kuster, F., & Orth, U. (2013). The long-term stability of self-esteem: its time-dependent decay and nonzero asymptote. *Personality and Social Psychology Bulletin*, *39*(5), 677–690. https://doi.org/10.1177/0146167213480189

Kuster, F., Orth, U., & Meier, L. L. (2013). High self-esteem prospectively predicts better work conditions and outcomes. *Social Psychological and Personality Science*, *4*(6), 668–675. https://doi.org/10.1177/1948550613479806

Lance, C. E., Christie, J., & Williamson, G. M. (2019). Do state and trait measures measure states and traits? The case of community-dwelling caregivers of older adults. *Assessment*, 1073191119888582. https://doi.org/10.1177/1073191119888582

Leary, M. R. (2004). The sociometer, self-esteem, and the resulation of interpersonal behavior. In R. F. Baumeister & K. D. Vohs (Eds.), *Handbook of self-regulation. Research, theory, and applications* (pp. 373–391). New York, NY: Guilford Press.

Leary, M. R., & Baumeister, R. F. (2000). The nature and function of self-esteem: Sociometer theory. In *Advances in Experimental Social Psychology. Advances in Experimental Social*

Psychology Volume 32 (Vol. 32, pp. 1–62). Elsevier. https://doi.org/10.1016/S0065-2601(00)80003-9

Lerner, R. M. (1998). Theories of human development: Contemporary perspectives. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Theoretical models of human development* (pp. 1–24). Hoboken, NJ: John Wiley & Sons, Inc.

Lerner, R. M. (2006). Developmental science, developmental systems, and contemporary theories of human development. In K. A. Renninger, I. E. Sigel, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology* (6th ed.). Hoboken: John Wiley & Sons. https://doi.org/10.1002/9780470147658.chpsy0101

Lerner, R. M., Lerner, J. V., Eye, A. von, Bowers, E. P., & Lewin-Bizan, S. (2011). Individual and contextual bases of thriving in adolescence: a view of the issues. *Journal of Adolescence*, *34*(6), 1107–1114. https://doi.org/10.1016/j.adolescence.2011.08.001

Lerner, R. M., Lerner, J. V., Lewin-Bizan, S., Bowers, E. P., Boyd, M. J., Mueller, M. K.,. . . Napolitano, C. M. (2011). Positive youth development: Processes, programs, and problematics. *Journal of Youth Development*, *6*(3), 38–62. https://doi.org/10.5195/JYD.2011.174

Lewin, K. (1939). Field theory and experiment in social psychology: Concepts and methods. *American Journal of Sociology*, *44*(6), 868–896. https://doi.org/10.1086/218177

Lewinsohn, P. M., Hoberman, H. M., & Rosenbaum, M. (1988). A prospective study of risk factors for unipolar depression. *Journal of Abnormal Psychology*, *97*(3), 251–264. https://doi.org/10.1037/0021-843X.97.3.251

Lewinsohn, P. M., Steinmetz, J. L., Larson, D. W., & Franklin, J. (1981). Depression-related cognitions: Antecedent or consequence? *Journal of Abnormal Psychology*, *90*(3), 213–219. https://doi.org/10.1037/0021-843X.90.3.213

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Lösch, T. (2016). *Perceiving achievement in schools: How do self-appraisals, peer appraisals and achievement relate to each other? (Doctoral dissertation, University of Tübingen)*. Retrieved from http://hdl.handle.net/10900/74162

Magnusson, D., & Stattin, H. (2006). The person in context: A holistic-interactionistic approach. In K. A. Renninger, I. E. Sigel, W. Damon, & R. M. Lerner (Eds.), *Handbook of*

child psychology (6th ed.). Hoboken: John Wiley & Sons. https://doi.org/10.1002/9780470147658.chpsy0108

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology*, *35*(2), 63–78. https://doi.org/10.1037/0022-3514.35.2.63

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, *23*(1), 129–149. https://doi.org/10.3102/00028312023001129

Marsh, H. W. (1987a). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, *79*(3), 280–295. https://doi.org/10.1037/0022-0663.79.3.280

Marsh, H. W. (1987b). The hierarchical structure of self-concept and the application of hierarchical confirmatory factor analysis. *Journal of Educational Measurement*, *24*(1), 17–39. https://doi.org/10.1111/j.1745-3984.1987.tb00259.x

Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology*, *82*(4), 623–636. https://doi.org/10.1037/0022-0663.82.4.623

Marsh, H. W. (1992). *Self Description Questionnaire (SDQ) II: A theoretical and empirical basis for the measurement of multiple dimensions of adolescent self-concept. A test manual and research monograph*. Macarthur, New South Wales, Australia: University of Western Sydney.

Marsh, H. W., Byrne, B. M., & Yeung, A. S. (1999). Causal ordering of academic self-concept and achievement: Reanalysis of a pioneering study and. *Educational Psychologist*, *34*(3), 155–167. https://doi.org/10.1207/s15326985ep3403_2

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

Marsh, H. W., Craven, R., & Martin, A. J. (2006). What is the nature of self-esteem? Unidimensional and multidimensional perspectives. In M. H. Kernis (Ed.), *Self-esteem issues and answers. A sourcebook of current perspectives.* New York: Psychology Press.

Marsh, H. W., & Hattie, J. A. (1996). Theoretical models in self-concept. In B. Bracken (Ed.), *Handbook of self-concept* (pp. 38–90). Hillsdale, NJ: Erlbaum.

Marsh, H. W., & Hau, K.-T. (2003). Big-fish-little-pond effect on academic self-concept. A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*, *58*(5), 364–376. https://doi.org/10.1037/0003-066x.58.5.364

Marsh, H. W., & Hau, K.-T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology*, *96*(1), 56–67. https://doi.org/10.1037/0022-0663.96.1.56

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*(1), 151–170. https://doi.org/10.1016/j.cedpsych.2006.10.008

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. *Psychological Bulletin*, *97*(3), 562–582. https://doi.org/10.1037/0033-2909.97.3.562

Marsh, H. W., Kuyper, H., Morin, A. J.S., Parker, P. D., & Seaton, M. (2014). Big-fish-little-pond social comparison and local dominance effects: Integrating new statistical models, methodology, design, theory and substantive implications. *Learning and Instruction*, *33*, 50–66. https://doi.org/10.1016/j.learninstruc.2014.04.002

Marsh, H. W., Liem, G. A. D., Martin, A. J., Morin, A. J. S., & Nagengast, B. (2011). Methodological Measurement Fruitfulness of Exploratory Structural Equation Modeling (ESEM): New Approaches to Key Substantive Issues in Motivation and Engagement. *Journal of Psychoeducational Assessment*, *29*(4), 322–346. https://doi.org/10.1177/0734282911406657

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, *44*(6), 764–802. https://doi.org/10.1080/00273170903333665

Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. S. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(2), 191–225. https://doi.org/10.1080/10705510902751010

Marsh, H. W., Martin, A. J., & Hau, K.-T. (2006). A multi method perspective on self-concept research in educational psychology: A construct validity approach. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology*. Washingtion, DC: American Psychological Association.

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 439–476. https://doi.org/10.1080/10705510903008220

Marsh, H. W., Nagengast, B., & Morin, A. J. S. (2013). Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. *Developmental Psychology*, *49*(6), 1194–1218. https://doi.org/10.1037/a0026913

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

Marsh, H. W., & O'Neill, R. (1984). Self Description Questionnaire III: The construct validity of multidimensional self-concept ratings by late adolescents. *Journal of Educational Measurement*, *21*(2), 153–174. https://doi.org/10.1111/j.1745-3984.1984.tb00227.x

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: A Multidisciplinary Journal*, *25*(1), 137–159. https://doi.org/10.1080/10705511.2017.1370377

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, *20*(3), 107–123. https://doi.org/10.1207/s15326985ep2003_1

Marsh, H. W., Trautwein, U., Lüdtke, O., & Köller, O. (2008). Social comparison and big-fish-little-pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology*, *100*(3), 510–524. https://doi.org/10.1037/0022-0663.100.3.510

Masselink, M., van Roekel, E., Hankin, B. L., Keijsers, L., Lodder, G.M.A., Vanhalst, J.,. . . Laceulle, O. (2018). The longitudinal association between self-esteem and depressive symptoms in adolescents: Separating between-person effects from within-person effects. *European Journal of Personality*, *96*, 35. https://doi.org/10.1002/per.2179

McCrae, R. R., & Costa, P. T. (2008). The five-factor theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality. Theory and research* (pp. 159–181). New York, NY: Guilford Press.

Mead, G. H. (1934). *Mind, self, and society from the standpoint of a social behaviorist*. Chicago, IL: University of Chicago Press.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. https://doi.org/10.1007/BF02294825

Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequence of measurement. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 33–45). Hillsdale, NJ: Erlbaum.

Möller, J., & Köller, O. (2001). Dimensional comparisons: An experimental approach to the internal/external frame of reference model. *Journal of Educational Psychology*, *93*(4), 826–835. https://doi.org/10.1037/0022-0663.93.4.826

Morin, A. J. S., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(1), 116–139. https://doi.org/10.1080/10705511.2014.961800

Morin, A. J. S., Maïano, C., Marsh, H. W., Janosz, M., & Nagengast, B. (2011). The longitudinal interplay of adolescents' self-esteem and body image: A conditional autoregressive latent trajectory analysis. *Multivariate Behavioral Research*, *46*(2), 157–201. https://doi.org/10.1080/00273171.2010.546731

Morin, A. J. S., Maïano, C., Marsh, H. W., Nagengast, B., & Janosz, M. (2013). School life and adolescents' self-esteem trajectories. *Child Development*, *84*(6), 1967–1988. https://doi.org/10.1111/cdev.12089

Mõttus, R., Kandler, C., Bleidorn, W., Riemann, R., & McCrae, R. R. (2017). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*, *112*(3), 474–490. https://doi.org/10.1037/pspp0000100

Mõttus, R., Sinick, J., Terracciano, A., Hřebíčková, M., Kandler, C., Ando, J.,. . . Jang, K. L. (2019). Personality characteristics below facets: A replication and meta-analysis of cross-rater agreement, rank-order stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*, *117*(4), e35-e50. https://doi.org/10.1037/pspp0000202

Mund, M., Finn, C., Hagemeyer, B., Zimmermann, J., & Neyer, F. J. (2015). The dynamics of self-esteem in partner relationships. *European Journal of Personality*, *29*(2), 235–249. https://doi.org/10.1002/per.1984

Murray, S. L., Holmes, J. G., & Collins, N. L. (2006). Optimizing assurance: the risk regulation system in relationships. *Psychological Bulletin*, *132*(5), 641–666. https://doi.org/10.1037/0033-2909.132.5.641

Murray, S. L., Holmes, J. G., & Griffin, D. W. (2000). Self-esteem and the quest for felt security: How perceived regard regulates attachment processes. *Journal of Personality and Social Psychology*, *78*(3), 478–498. https://doi.org/10.1037//0022-3514.78.3.478

Nagengast, B., & Marsh, H. W. (2011). The negative effect of school-average ability on science self-concept in the UK, the UK countries and the world: the Big-Fish-Little-Pond-Effect for PISA 2006. *Educational Psychology*, *31*(5), 629–656. https://doi.org/10.1080/01443410.2011.586416

Nagengast, B., Marsh, H. W., Scalas, L. F., Xu, M. K., Hau, K.-T., & Trautwein, U. (2011). Who took the “x” out of expectancy-value theory? A psychological mystery, a substantive-methodological synergy, and a cross-national generalization. *Psychological Science*, *22*(8), 1058–1066. https://doi.org/10.1177/0956797611415540

Nezlek, J. B. (2007). A multilevel framework for understanding relationships among traits, states, situations and behaviours. *European Journal of Personality*, *21*(6), 789–810. https://doi.org/10.1002/per.640

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

OECD. (2017a). *Education at a glance 2017: OECD indicators*. Paris: OECD Publishing.

OECD. (2017b). *PISA 2015 results (Volume III): Students' well-being*. Paris: OECD Publishing.

O'Mara, A. J., Marsh, H. W., Craven, R. G., & Debus, R. L. (2006). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist*, *41*(3), 181–206. https://doi.org/10.1207/s15326985ep4103_4

Orth, U., Clark, D. A., Donnellan, M. B., & Robins, R. W. (2020). Testing prospective effects in longitudinal research: Comparing seven competing cross-lagged models. *Journal of Personality and Social Psychology*.

Orth, U. (2018). The family environment in early childhood has a long-term effect on self-esteem: A longitudinal study from birth to age 27 years. *Journal of Personality and Social Psychology*, *114*(4), 637–655. https://doi.org/10.1037/pspp0000143

Orth, U., Erol, R. Y., & Luciano, E. C. (2018). Development of self-esteem from age 4 to 94 years: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *144*(10), 1045–1080. https://doi.org/10.1037/bul0000161

Orth, U., & Luciano, E. C. (2015). Self-esteem, narcissism, and stressful life events: Testing for selection and socialization. *Journal of Personality and Social Psychology*, *109*(4), 707–721. https://doi.org/10.1037/pspp0000049

Orth, U., Maes, J., & Schmitt, M. (2015). Self-esteem development across the life span: a longitudinal study with a large sample from Germany. *Developmental Psychology*, *51*(2), 248–259. https://doi.org/10.1037/a0038481

Orth, U., & Robins, R. W. (2013). Understanding the link between low self-esteem and depression. *Current Directions in Psychological Science*, *22*(6), 455–460. https://doi.org/10.1177/0963721413492763

Orth, U., & Robins, R. W. (2014). The development of self-esteem. *Current Directions in Psychological Science*, *23*(5), 381–387. https://doi.org/10.1177/0963721414547414

Orth, U., & Robins, R. W. (2019). Development of self-esteem across the lifespan. In D. P. McAdams, R. L. Shiner, & J. L. Tackett (Eds.), *Handbook of personality development* (pp. 328–344). New York, NY: The Guilford Press.

Orth, U., Robins, R. W., & Meier, L. L. (2009). Disentangling the effects of low self-esteem and stressful events on depression: Findings from three longitudinal studies. *Journal of Personality and Social Psychology*, *97*(2), 307–321. https://doi.org/10.1037/a0015645

Orth, U., Robins, R. W., & Roberts, B. W. (2008). Low self-esteem prospectively predicts depression in adolescence and young adulthood. *Journal of Personality and Social Psychology*, *95*(3), 695–708. https://doi.org/10.1037/0022-3514.95.3.695

Orth, U., Robins, R. W., & Widaman, K. F. (2012). Life-span development of self-esteem and its effects on important life outcomes. *Journal of Personality and Social Psychology*, *102*(6), 1271–1288. https://doi.org/10.1037/a0025558

Orth, U., Robins, R. W., Widaman, K. F., & Conger, R. D. (2014). Is low self-esteem a risk factor for depression? Findings from a longitudinal study of Mexican-origin youth. *Developmental Psychology*, *50*(2), 622–633. https://doi.org/10.1037/a0033817

Orth, U., Trzesniewski, K. H., & Robins, R. W. (2010). Self-esteem development from young adulthood to old age: a cohort-sequential longitudinal study. *Journal of Personality and Social Psychology*, *98*(4), 645–658. https://doi.org/10.1037/a0018769

Piaget, J. (1960). *The psychology of intelligence*. Patterson, NJ: Littlefield, Adams.

Podsakoff, N. P., Spoelma, T. M., Chawla, N., & Gabriel, A. S. (2019). What predicts within-person variance in applied psychology constructs? An empirical examination. *The Journal of Applied Psychology*, *104*(6), 727–754. https://doi.org/10.1037/apl0000374

Ravens-Sieberer, U., Gosch, A., Abel, T., Auquier, P., Bellach, B.-M., Bruil, J.,. . . Rajmil, L. (2001). Quality of life in children and adolescents: A European public health perspective. *Sozial- und Präventivmedizin SPM*, *46*(5), 294–302. https://doi.org/10.1007/BF01321080

Reitz, A. K., Motti-Stefanidi, F., & Asendorpf, J. B. (2016). Me, us, and them: Testing sociometer theory in a socially diverse real-life context. *Journal of Personality and Social Psychology*, *110*(6), 908–920. https://doi.org/10.1037/pspp0000073

Rentzsch, K., & Schröder-Abé, M. (2018). Stability and change in domain-specific self-esteem and global self-esteem. *European Journal of Personality*, *32*(4), 353–370. https://doi.org/10.1002/per.2167

Rentzsch, K., Wenzler, M. P., & Schütz, A. (2016). The structure of multidimensional self-esteem across age and gender. *Personality and Individual Differences*, *88*, 139–147. https://doi.org/10.1016/j.paid.2015.09.012

Rhemtulla, M., van Bork, R., & Borsboom, D. (2015). Calling models with causal indicators "measurement models" implies more than they can deliver. *Measurement: Interdisciplinary Research and Perspectives*, *13*(1), 59–62. https://doi.org/10.1080/15366367.2015.1016343

Rhemtulla, M., van Bork, R., & Borsboom, D. (2019). Worse than measurement error: Consequences of inappropriate latent variable measurement models. *Psychological Methods.* Advance online publication. https://doi.org/10.1037/met0000220

Rieger, S., Göllner, R., Spengler, M., Trautwein, U., Nagengast, B., & Roberts, B. W. (2017). Social cognitive constructs are just as stable as the big five between grades 5 and 8. *AERA Open*, *3*(3), 233285841771769. https://doi.org/10.1177/2332858417717691

Rieger, S., Göllner, R., Trautwein, U., & Roberts, B. W. (2016). Low self-esteem prospectively predicts depression in the transition to young adulthood: A replication of Orth, Robins, and Roberts (2008). *Journal of Personality and Social Psychology*, *110*(1), e16-e22. https://doi.org/10.1037/pspp0000037

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. *Psychological Bulletin*, *126*(1), 3–25. https://doi.org/10.1037/0033-2909.126.1.3

Roberts, B. W., & Wood, D. (2006). Personality development in the context of the neo-socioanalytic model of personality. In D. K. Mroczek & T. D. Little (Eds.), *Handbook of persoanlity development* (pp. 11–39). Lawrence Erlbaum Associates Publishers.

Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. *Journal of Research in Personality*, *43*(2), 137–145. https://doi.org/10.1016/j.jrp.2008.12.015

Roberts, B. W. (2018). A revised sociogenomic model of personality traits. *Journal of Personality*, *86*, 23–35. https://doi.org/10.1111/jopy.12323

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies. *Psychological Bulletin*, *132*(1), 1–25. https://doi.org/10.1037/0033-2909.132.1.1

Robins, J. M., Hernán, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology (Cambridge, Mass.)*, *11*(5), 550–560. https://doi.org/10.1097/00001648-200009000-00011

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, *27*(2), 151–161. https://doi.org/10.1177/0146167201272002

Robins, R. W., Tracy, J. L., Trzesniewski, K., Potter, J., & Gosling, S. D. (2001). Personality correlates of self-esteem. *Journal of Research in Personality*, *35*(4), 463–482. https://doi.org/10.1006/jrpe.2001.2324

Robinson, M. D., & Clore, G. L. (2002). Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes. *Journal of Personality and Social Psychology*, *83*(1), 198–215. https://doi.org/10.1037/0022-3514.83.1.198

Roeser, R. W., Eccles, J. S., & Sameroff, A. J. (2000). School as a context of early adolescents' academic and social-emotional development: A summary of research findings. *The Elementary School Journal*, *100*(5), 443–471. https://doi.org/10.1086/499650

Rogosa, D. (1980). A critique of cross-lagged correlation. *Psychological Bulletin*, *88*(2), 245–258. https://doi.org/10.1037/0033-2909.88.2.245

Rose, N., Wagner, W., Mayer, A., & Nagengast, B. (2019). Model-based manifest and latent composite scores in structural equation models. *Collabra: Psychology*, *5*(1), 9. https://doi.org/10.1525/collabra.143

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rosenberg, M. (1979). *Conceiving the self*. New York: Basic Books.

Rosenberg, M. (1986). Self-concept from middle childhood through adolescence. In J. Suls & A. G. Greenwald (Eds.), *Psychological perspectives on the self* (pp. 107–136). Hillsdale, NJ: Erlbaum Press.

Rosenberg, M. (1989). *Society and the adolescent self-image*. Middletown, CT: Wesleyan University Press.

Rosenberg, M., Schooler, C., Schoenbach, C., & Rosenberg, F. (1995). Global self-esteem and specific self-esteem: Different concepts, different outcomes. *American Sociological Review, 60*(1), 141. https://doi.org/10.2307/2096350

Ryan, R. M., Stiller, J. D., & Lynch, J. H. (1994). Representations of relationships to teachers, parents, and friends as predictors of academic motivation and self-esteem. *The Journal of Early Adolescence, 14*(2), 226–249. https://doi.org/10.1177/027243169401400207

Sameroff, A. J. (2009). *The transactional model of development: How children and contexts shape each other* (1st ed.). Washington, D.C: American Psychological Association. Retrieved from http://search.ebscohost.com/direct.asp?db=pzh&jid=%22200905801%22&scope=site

Schaffhuser, K., Wagner, J., Lüdtke, O., & Allemand, M. (2014). Dyadic longitudinal interplay between personality and relationship satisfaction: A focus on neuroticism and self-esteem. *Journal of Research in Personality, 53*, 124–133. https://doi.org/10.1016/j.jrp.2014.08.007

Scherrer, V., & Preckel, F. (2019). Development of motivational variables and self-esteem during the school career: A meta-analysis of longitudinal studies. *Review of Educational Research*, *89*(2), 211–258. https://doi.org/10.3102/0034654318819127

Schmitt, D. P., & Allik, J. (2005). Simultaneous administration of the Rosenberg Self-Esteem Scale in 53 nations: exploring the universal and culture-specific features of global self-esteem. *Journal of Personality and Social Psychology*, *89*(4), 623–642. https://doi.org/10.1037/0022-3514.89.4.623

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Shahar, G., & Davidson, L. (2003). Depressive symptoms erode self-esteem in severe mental illness: A three-wave, cross-lagged study. *Journal of Consulting and Clinical Psychology, 71*(5), 890–900. https://doi.org/10.1037/0022-006X.71.5.890

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, *46*(3), 407–441. https://doi.org/10.3102/00346543046003407

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

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

Smith, G. T. (2005). On construct validity: Issues of method and measurement. *Psychological Assessment*, *17*(4), 396–408. https://doi.org/10.1037/1040-3590.17.4.396

Soest, T. von, Wichstrøm, L., & Kvalem, I. L. (2016). The development of global and domain-specific self-esteem from age 13 to 31. *Journal of Personality and Social Psychology*, *110*(4), 592–608. https://doi.org/10.1037/pspp0000060

Soto, C. J., & Tackett, J. L. (2015). Personality traits in childhood and adolescence. *Current Directions in Psychological Science*, *24*(5), 358–362. https://doi.org/10.1177/0963721415589345

Sowislo, J. F., & Orth, U. (2013). Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies. *Psychological Bulletin*, *139*(1), 213–240. https://doi.org/10.1037/a0028931

Spielberger, C. D. (1966). *Anxiety and behavior*. New York: Academic Press.

Spielberger, C. D. (1995). *State-Trait Depression Scales (Form X-1)*. Palo Alto, CA: Mind Garden.

Srivastava, S., & Beer, J. S. (2005). How self-evaluations relate to being liked by others: Integrating sociometer and attachment perspectives. *Journal of Personality and Social Psychology*, *89*(6), 966–977. https://doi.org/10.1037/0022-3514.89.6.966

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Steyer, R., Ferring, D., & Schmitt, M. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment*. (8), 79–98.

Steyer, R., & Schmitt, T. (1994). The theory of confounding and its application in causal modeling with latent variables. In A. von Eye & C. C. Clogg (Eds.), *Latent variable analysis: Applications for developmental research* (pp. 36–67). Thousand Oaks, CA: Sage.

Steyer, R., Mayer, A., Geiser, C., & Cole, D. A. (2015). A theory of states and traits-revised. *Annual Review of Clinical Psychology*, *11*, 71–98. https://doi.org/10.1146/annurev-clinpsy-032813-153719

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality*, *13*(5), 389–408. https://doi.org/10.1002/(SICI)1099-0984(199909/10)13:5<389::AID-PER361>3.0.CO;2-A

Swann, W. B. (1996). *Self-traps: The elusive quest for higher self-esteem* (1. print). New York: Freeman.

Swann, W. B., JR, Chang-Schneider, C., & Larsen McClarty, K. (2007). Do people's self-views matter? Self-concept and self-esteem in everyday life. *American Psychologist*, *62*(2), 84–94. https://doi.org/10.1037/0003-066X.62.2.84

Tafarodi, R.W., & Swann, W.B. (2001). Two-dimensional self-esteem: theory and measurement. *Personality and Individual Differences*, *31*(5), 653–673. https://doi.org/10.1016/S0191-8869(00)00169-0

Tafarodi, R. W., & Swann, W. B. (1996). Individualism-Collectivism and Global Self-Esteem. *Journal of Cross-Cultural Psychology*, *27*(6), 651–672. https://doi.org/10.1177/0022022196276001

Tesser, A., Crepaz, N., Collins, J. C., Cornell, D., & Beach, S. R. H. (2000). Confluence of Self-Esteem Regulation Mechanisms: On Integrating the Self-Zoo. *Personality and Social Psychology Bulletin*, *26*(12), 1476–1489. https://doi.org/10.1177/01461672002612003

Tesser, A., Millar, M., & Moore, J. (1988). Some affective consequences of social comparison and reflection processes: The pain and pleasure of being close. *Journal of Personality and Social Psychology*, *54*(1), 49–61. https://doi.org/10.1037/0022-3514.54.1.49

Tetzner, J., Becker, M., & Baumert, J. (2016). Still Doing Fine? The Interplay of Negative Life Events and Self-Esteem During Young Adulthood. *European Journal of Personality*, *30*(4), 358–373. https://doi.org/10.1002/per.2066

Trautwein, U. (2003). *Schule und Selbstwert [Schools and self-esteem]*. Münster: Waxmann.

Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Self-esteem, academic self-concept, and achievement: how the learning environment moderates the dynamics of self-concept. *Journal of Personality and Social Psychology*, *90*(2), 334–349. https://doi.org/10.1037/0022-3514.90.2.334

Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2013). Development of self-esteem. In V. Zeigler-Hill (Ed.), *Self-esteem* (pp. 60–79). London, UK: Psychology Press.

Trzesniewski, K. H., Donnellan, M. B., Moffitt, T. E., Robins, R. W., Poulton, R., & Caspi, A. (2006). Low self-esteem during adolescence predicts poor health, criminal behavior, and limited economic prospects during adulthood. *Developmental Psychology*, *42*(2), 381–390. https://doi.org/10.1037/0012-1649.42.2.381

Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2003). Stability of self-esteem across the life span. *Journal of Personality and Social Psychology*, *84*(1), 205–220.

Twenge, J. M., Carter, N. T., & Campbell, W. K. (2017). Age, time period, and birth cohort differences in self-esteem: Reexamining a cohort-sequential longitudinal study. *Journal of Personality and Social Psychology*, *112*(5), e9-e17. https://doi.org/10.1037/pspp0000122

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

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*(9), e0209133. https://doi.org/10.1371/journal.pone.0209133

van ‘t Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology*, *67*, 2–12. https://doi.org/10.1016/j.jesp.2016.03.004

Vazire, S., & Sherman, R. A. (2017). Introduction to the special issue on within-person variability in personality. *Journal of Research in Personality*, *69*, 1–3. https://doi.org/10.1016/j.jrp.2017.07.004

Vernon, P. E. (1950). *The structure of human abilities*. New York: Wiley.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*(6), 632–638. https://doi.org/10.1177/1745691612463078

Wagner, J., Lüdtke, O., Jonkmann, K., & Trautwein, U. (2013). Cherish yourself: longitudinal patterns and conditions of self-esteem change in the transition to young adulthood. *Journal of Personality and Social Psychology*, *104*(1), 148–163. https://doi.org/10.1037/a0029680

Wagner, J., Lüdtke, O., Robitzsch, A., Göllner, R., & Trautwein, U. (2018). Self-esteem development in the school context: The roles of intrapersonal and interpersonal social predictors. *Journal of Personality*, *86*(3), 481–497. https://doi.org/10.1111/jopy.12330

Wagner, J., Lüdtke, O., & Trautwein, U. (2016). Self-esteem is mostly stable across young adulthood: Evidence from Latent STARTS Models. *Journal of Personality*, *84*(4), 523–535. https://doi.org/10.1111/jopy.12178

Wagner, J., Orth, U., Bleidorn, W., Hopwood, C. J., & Kandler, C. (in press). Towards an integrative model of sources of personality stability and change. *Perspectives on Psychological Science.* Retrieved from https://doi.org/10.31234/osf.io/qzef8

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin*, *83*(2), 213–217. https://doi.org/10.1037/0033-2909.83.2.213

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Watson, D., Suls, J., & Haig, J. (2002). Global self-esteem in relation to structural models of personality and affectivity. *Journal of Personality and Social Psychology*, *83*(1), 185–197. https://doi.org/10.1037/0022-3514.83.1.185

Webster, G. D., Smith, C. V., Brunell, A. B., Paddock, E. L., & Nezlek, J. B. (2017). Can Rosenberg's (1965) Stability of Self Scale capture within-person self-esteem variability? Meta-analytic validity and test–retest reliability. *Journal of Research in Personality*, *69*, 156–169. https://doi.org/10.1016/j.jrp.2016.06.005

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, *4*(1), 10–18. https://doi.org/10.1111/j.1750-8606.2009.00110.x

Wigfield, A., & Eccles, J. S. (1994). Children's competence beliefs, achievement values, and general self-esteem. *The Journal of Early Adolescence*, *14*(2), 107–138. https://doi.org/10.1177/027243169401400203

Wittmann, W. W. (1988). Multivariate reliability theory: Principles of symmetry and successful validation strategies. In J. R. Nesselroade & R. B. Catell (Eds.), *Handbook of multivariate experimental psychology* (pp. 505–560). New York, NY: Plenum Press.

World Health Organization. (2008). *The global burden of disease: 2004 update*. Geneva, Switzerland: Author.

Wrzus, C., & Roberts, B. W. (2017). Processes of personality development in adulthood: The TESSERA framework. *Personality and Social Psychology Review*, *21*(3), 253–277. https://doi.org/10.1177/1088868316652279

Zeigler-Hill, V. (Ed.). (2013). *Self-esteem*. London, UK: Psychology Press.

Zuckerman, M., Li, C., & Hall, J. A. (2016). When men and women differ in self-esteem and when they don't: A meta-analysis. *Journal of Research in Personality*, *64*, 34–51. https://doi.org/10.1016/j.jrp.2016.07.007

Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z.,. . . Diener, E. (2019). From data to causes I: Building a general cross-lagged panel model (GCLM). *Organizational Research Methods*, *175*(4), 109442811984727. https://doi.org/10.1177/1094428119847278

Zyphur, M. J., Voelkle, M. C., Tay, L., Allison, P. D., Preacher, K. J., Zhang, Z.,. . . Diener, E. (2019). From data to causes II: Comparing approaches to panel data analysis. *Organizational Research Methods*, *101*(2), 109442811984728. https://doi.org/10.1177/1094428119847280