Development of Domain-Specific Self-Evaluations:

A Meta-Analysis of Longitudinal Studies

Ulrich Orth, Laura C. Dapp, Ruth Yasemin Erol, Samantha Krauss, and Eva C. Luciano

University of Bern

Orth, U., Dapp, L. C., Erol, R. Y., Krauss, S., & Luciano, E. C. (2021). Development of domain-specific self-evaluations: A meta-analysis of longitudinal studies. *Journal of Personality and Social Psychology*, *120*, 145-172. https://doi.org/10.1037/pspp0000378

Author Note

Ulrich Orth, Laura C. Dapp, Ruth Yasemin Erol, Samantha Krauss, and Eva C. Luciano, Department of Psychology, University of Bern.

Data, materials, and code are available on the Open Science Framework (https://osf.io/hnrv5/).

Correspondence concerning this article should be addressed to Ulrich Orth, Department of Psychology, University of Bern, Fabrikstrasse 8, 3012 Bern, Switzerland. E-mail: ulrich.orth@psy.unibe.ch

Abstract

This meta-analysis investigated the normative development of domain-specific self-evaluations (also referred to as self-concept or domain-specific self-esteem) by synthesizing the available longitudinal data on mean-level change. Eight domains of self-evaluations were assessed: academic abilities, athletic abilities, physical appearance, morality, romantic relationships, social acceptance, mathematics, and verbal abilities. Analyses were based on data from 143 independent samples which included 112,204 participants. As the effect size measure, we used the standardized mean change *d* per year. The mean age associated with effect sizes ranged from 5 to 28 years. Overall, developmental trajectories of self-evaluations were positive in the domains of academic abilities, social acceptance, and romantic relationships. In contrast, self-evaluations showed negative developmental trajectories in the domains of morality, mathematics, and verbal abilities. Little mean-level change was observed for self-evaluations of physical appearance and athletic abilities. Moderator analyses were conducted for the full set of samples and for the subset of samples between ages 10 and 16 years. The moderator analyses indicated that the pattern of findings held across demographic characteristics of the samples, including gender and birth cohort. The meta-analytic dataset consisted largely of Western and White/European samples, pointing to the need of conducting more research with Non-Western and ethnically diverse samples. The meta-analytic findings suggest that the notion that self-evaluations generally show a substantial decline in the transition from early to middle childhood should be revised. Also, the findings did not support the notion that self-evaluations reach a critical low point in many domains in early adolescence.

Keywords: domain-specific self-evaluations, self-esteem, development, longitudinal studies, meta-analysis

Development of Domain-Specific Self-Evaluations:

A Meta-Analysis of Longitudinal Studies

When people think about whether they are satisfied with their competences and personal characteristics, they may come to very different conclusions. Some individuals may be pleased with their competences in some domains (e.g., school or work), but dissatisfied with other domains (e.g., attractiveness). Others may feel that they do reasonably well with regard to many criteria, but that there is no single domain in which they really excel. Still others may have very positive self-views across the board. Notwithstanding these interindividual differences, it is possible that the average positivity of self-evaluations follows a systematic, normative trajectory across childhood, adolescence, and adulthood. Moreover, the normative developmental trajectory might be domain-specific; that is, the trajectory might depend on whether self-evaluations refer to the academic domain, social relationships, sports, or physical appearance.

Many studies examined developmental trajectories of domain-specific self-evaluations, both decades ago (e.g., Cairns, McWhirter, Duffy, & Barry, 1990; Cole et al., 2001; Eccles et al., 1989; Marsh, 1989) and more recently (e.g., Esnaola, Sesé, Antonio-Agirre, & Azpiazu, 2020; Harris, Wetzel, Robins, Donnellan, & Trzesniewski, 2018; Kuzucu, Bontempo, Hofer, Stallings, & Piccinin, 2014). However, despite the large body of evidence that has been accumulated, the findings have been inconsistent and have not yet led to any agreement on the normative development of domain-specific self-evaluations. In fact, the inconsistent pattern of findings in the literature has been emphasized by the authors of many previous studies on the topic (e.g., Cole et al., 2001; Harris et al., 2018; Kuzucu et al., 2014; von Soest, Wichstrom, & Kvalem, 2016). In this situation, meta-analytic methods are ideally suited to gain more robust insights into developmental patterns, by aggregating the evidence across a large number of studies. Therefore,

the goal of the present research was to synthesize the available data on how and whether domain-specific self-evaluations change as a function of age.

Understanding the development of domain-specific self-evaluations is important because research suggests that self-evaluations prospectively predict outcomes in important life domains, such as education, work, and health (Marsh & Craven, 2006; Spilt, van Lier, Leflot, Onghena, & Colpin, 2014; Steiger, Allemand, Robins, & Fend, 2014; Valentine, DuBois, & Cooper, 2004; von Soest et al., 2016). For example, favorable self-evaluations in the domain of academic abilities predict better academic achievement, when prior levels of achievement are controlled for (Marsh & Craven, 2006; Marsh & Yeung, 1997). Also, understanding the development of domain-specific self-evaluations may contribute to a better understanding of the development of global self-esteem (Orth & Robins, 2019).

In this article, we use the term domain-specific self-evaluations, rather than self-concept or domain-specific self-esteem, to denote children's and adults' self-perceptions in domains such as academic abilities, athletic abilities, physical appearance, and social acceptance. We decided to use the term self-evaluations because the relevant measures assess the individual's evaluative beliefs about their abilities or other favorable characteristics, that is, self-evaluations on a continuous dimension ranging from bad (or low, weak) to good (or high, strong). In contrast, the term self-concept also includes non-evaluative beliefs about the self and the term self-esteem is often understood as a person's general level of self-acceptance. However, it is important to note that domain-specific self-evaluations, self-concept, and domain-specific self-esteem are established terms for the same construct, so in the literature summarized in this meta-analysis, all three terms are often used synonymously.

Theoretical Perspectives on the Development of Domain-Specific Self-Evaluations

In the following sections, we review theoretical perspectives on normative change in domain-specific self-evaluations from childhood to adulthood, on similarity to normative change in global self-esteem, and on moderators of change in domain-specific self-evaluations.

Declines in the Transition From Early to Middle Childhood

Theory suggests that average levels of self-evaluations decline in the transition from early to middle childhood (i.e., from about age 5 to 8 years), resulting from a number of social-cognitive advances (Harter, 2006b, 2006c; Robins, Trzesniewski, Tracy, Gosling, & Potter, 2002). For example, children at age 4 or 5 years view their self in an overly positive light because they cannot yet discriminate between their actual and their ideal self (Harter, 2006b). In later years, when children learn to distinguish between actual and ideal characteristics, this process leads to less positive self-evaluations in most children. Another social-cognitive advance in this period is that children learn to use social comparison information when evaluating themselves (Gore & Cross, 2014; Harter, 2006b; Ruble, Boggiano, Feldman, & Loebl, 1980). At age 4 or 5 years, children predominantly make temporal comparisons (i.e., comparing their present competences with their competences a few months earlier, typically leading to positive self-evaluation; Harter, 2006c). However, after entering primary school—a transition that occurs in many countries at about age 6 years—they begin to compare their competences with those of their classmates and peers, which leads to reduced positivity in self-evaluations in most children. Another social-cognitive change emphasized by Harter (2006a) is the improvement in perspective-taking skills. During the transition from early to middle childhood, children typically make substantial progress in social perspective taking, which allows them to infer how the self is evaluated by others (e.g., peers, parents, and teachers). Again, this process causes gradual negative changes in self-evaluations in this period, because children learn, for example, that not

all classmates view their social, athletic, and academic competences positively. In addition to these social-cognitive changes, Harter (2006a) stresses that most parents increase their expectations for their children in this period, for example with regard to behavioral conduct and social competences. If children internalize these raised expectations in their self-concept, this leads to further declines in their self-evaluations.

In sum, the literature suggests that mean levels of self-evaluations decline in the transition from early to middle childhood, in particular because social-cognitive advances reduce the unrealistically positive self-perceptions that are characteristic of very young children at age 4 or 5 years (Harter, 2006c). The literature does not advance hypotheses that are specific to different domains of self-evaluations. Rather, the theoretical perspectives described in this section suggest that the decline applies to all domains of self-evaluation.

Low Point in Early Adolescence

Theory also suggests that individuals typically experience a low point in their self-evaluations in early adolescence (i.e., from about age 10 to 14 years), resulting from developmental changes related to puberty and to the transition from primary to secondary school (Harter, 2006c). Puberty could challenge adolescents' self-evaluations because of the many biological, cognitive, emotional, and social changes in this period (Harter, 2006c; Simmons, Blyth, Van Cleave, & Bush, 1979; Wigfield, Eccles, Mac Iver, Reuman, & Midgley, 1991). In addition, the transition from primary to secondary school could lead to normative declines in self-evaluations because it involves stricter grading standards and greater emphasis on evaluation, less personal attention by teachers, greater social comparison among adolescents, and a disruption in the adolescent's social network (Becker & Neumann, 2018; Cantin & Boivin, 2004; Eccles, Midgley, & Adler, 1984; Fenzel, 2000; Gniewosz, Eccles, & Noack, 2012;

Wigfield et al., 1991). In fact, several studies suggest that self-evaluations become more negative in early adolescence, in particular in the domains of academic abilities and physical appearance (Eccles et al., 1984; Marsh, 1989; Wigfield et al., 1991). The hypothesis that self-evaluations are relatively low in early adolescence is also consistent with research on personality development, which suggests that adaptive traits such as conscientiousness, agreeableness, and emotional stability decline and reach their low points in early adolescence (Göllner et al., 2017; Soto, 2016; Soto, John, Gosling, & Potter, 2011; Soto & Tackett, 2015).

In sum, even if early adolescence is no longer considered a period of inevitable storm and stress (Arnett, 1999; Hollenstein & Lougheed, 2013), the literature suggests that mean levels of self-evaluations reach a nadir in early adolescence. The perspectives reviewed above suggest that the low point might be particularly pronounced with regard to self-evaluations of academic abilities, social acceptance, and physical appearance.

Increases in Late Adolescence and Young Adulthood

The literature proposes that mean levels of self-evaluations begin to increase in late adolescence (i.e., from about age 16 years) and continue to increase in young adulthood, because in these developmental periods most personality characteristics develop in the direction of greater maturity (Harter, 2006b). In fact, a large number of studies suggest that people typically become more conscientious (including behaviors such as delaying gratification, following rules, and planning), agreeable (showing more prosocial behavior, trust, and modesty), and emotionally stable (being more even-tempered and experiencing less negative affect) as they move through late adolescence and young adulthood (Lucas & Donnellan, 2011; Roberts, Walton, & Viechtbauer, 2006; Roberts, Wood, & Caspi, 2008; Terracciano, McCrae, Brant, & Costa, 2005), a pattern that has been labeled the maturity principle of personality development (Roberts &

Wood, 2006; Roberts et al., 2008). Given that conscientiousness, agreeableness, and emotional stability contribute to adaptive functioning in social relationships, educational contexts, and the work domain (Mund & Neyer, 2014; Neyer & Asendorpf, 2001; Noftle & Robins, 2007; Roberts, Caspi, & Moffitt, 2003), the maturity principle suggests that people's social, academic, and job-related competences increase as a consequence of personality development. Ultimately, then, objective improvements in these competences could lead to more positive self-evaluations in these domains. Moreover, conscientiousness is also a key predictor of health behavior and exercising (Shanahan, Hill, Roberts, Eccles, & Friedman, 2014). Again, the normative increase in conscientiousness during late adolescence and young adulthood could lead to more positive self-evaluations with regard to athletic abilities and physical appearance, mediated by improved behavior with regard to exercise, diet, sleep, and personal hygiene. An additional reason for increasing levels of self-evaluations is related to possible changes in the relevant reference groups. During late adolescence and young adulthood, individuals gain more autonomy and control in selecting social, educational, and work contexts that match their personality, interests, and abilities (Harter, 2006c; Roberts et al., 2008). Thus, individuals may seek and enter environments in which their competences are more appreciated than in other environments (i.e., niche picking). The corresponding changes in the person's reference groups may lead to favorable changes in self-evaluations (Harter, 2006b).

In sum, there is reason to expect that mean levels of domain-specific self-evaluations increase in late adolescence and young adulthood. The literature on personality development suggests that these improvements might be particularly pronounced with regard to domains that are influenced by maturation of personality, such as academic abilities, social acceptance, and romantic relationships.

Similarity to the Trajectory of Global Self-Esteem

The literature also suggests that in some domains self-evaluations show trajectories that are similar to the trajectory of global self-esteem, whereas in other domains the trajectories differ from it more strongly (e.g., Harris et al., 2018; Marsh, 1989; von Soest et al., 2016). Over the past decade, many longitudinal studies have focused on the development of global self-esteem (e.g., Chung et al., 2014; Orth, Trzesniewski, & Robins, 2010; Wagner, Lüdtke, Jonkmann, & Trautwein, 2013; for a review, see Orth & Robins, 2019). A recent meta-analysis suggested that global self-esteem, on average, increases in early and middle childhood, remains stable in adolescence, increases strongly in young and middle adulthood, and declines in old age (Orth, Erol, & Luciano, 2018). This pattern of findings held across gender, country, ethnicity, and birth cohort.

There is reason to expect that domain-specific self-evaluations that correlate more strongly (versus less strongly) with global self-esteem show developmental trajectories that are more similar (versus less similar) to the trajectory of global self-esteem. Although all domain-specific self-evaluations typically show positive correlations with global self-esteem, the domains differ in the strength of their association with it (Donnellan, Trzesniewski, Conger, & Conger, 2007; Esnaola et al., 2020; Harris et al., 2018; Marsh, 1986; von Soest et al., 2016). Across studies, the strongest correlations emerged for self-evaluations in the domains of physical appearance, academic abilities, and social acceptance. In contrast, lower correlations are typically reported for self-evaluations of athletic abilities, behavioral conduct, and specific academic subjects such as mathematics. These correlational findings raise the possibility that self-evaluations in the appearance, academic, and social domain show average trajectories that

are relatively similar to the trajectory of global self-esteem, whereas the average trajectories of other domain-specific measures differ more strongly from global self-esteem.

Sociometer theory (Leary, 2012; Leary & Baumeister, 2000) suggests that self-evaluations in the social domain should show mean-level trends that are especially similar to global self-esteem. More precisely, this theory proposes that global self-esteem reflects an individual's relational value (i.e., with regard to inclusion in close relationships and small groups), as subjectively perceived by the individual him- or herself. In fact, research has documented strong associations between global self-esteem and inclusion in social relationships (Denissen, Penke, Schmitt, & Van Aken, 2008; Leary, Haupt, Strausser, & Chokel, 1998; Murray, Griffin, Rose, & Bellavia, 2003), supporting this central proposition of sociometer theory. Thus, this perspective suggests that the normative trajectory of self-evaluations with regard to social acceptance and romantic relationships could match the trajectory of global self-esteem relatively closely.

Moderators of Mean-Level Change in Domain-Specific Self-Evaluations

Finally, an important theme in the literature is the search for moderators of the trajectories of domain-specific self-evaluations. Many studies have examined the role of gender. In fact, meta-analyses have documented cross-sectional gender differences that largely correspond to gender stereotypes about self-evaluations (Gentile et al., 2009; Wilgenbusch & Merrell, 1999). The meta-analytic findings suggest that girls and women show more positive self-evaluations with regard to morality and verbal abilities, whereas boys and men show more positive self-evaluations with regard to athletic abilities, physical appearance, and mathematics (all of these effects were of small to medium size, i.e., absolute *d* values ranged from about .20 to .40). With regard to self-evaluations of general academic abilities and social acceptance, the

gender differences were nonsignificant and close to zero (Gentile et al., 2009; Wilgenbusch & Merrell, 1999). However, although there is robust evidence for cross-sectional gender differences in domain-specific self-evaluations (i.e., in the average level of self-evaluations), the evidence on gender differences in mean-level change (i.e., in the average slope of trajectories) is less clear. For example, in the study by von Soest et al. (2016), gender differences in the intercepts of trajectories largely corresponded to the meta-analytic findings described above, but gender differences in the slopes of trajectories were much closer to zero and mostly nonsignificant (for similar results see Cole et al., 2001; Kuzucu et al., 2014; Marsh, 1989). Thus, the literature does not advance hypotheses about gender differences in normative *change* in specific domains.

Birth cohort membership is another factor that might explain individual differences in the trajectory of self-evaluations. Some studies have suggested that more recent generations show greater positivity in their self-concept and even unrealistically positive self-evaluations (Gentile, Twenge, & Campbell, 2010; Twenge & Campbell, 2001, 2008; Twenge, Konrath, Foster, Campbell, & Bushman, 2008). If so, then the normative trajectories of self-evaluations in members of more recent generations might be characterized by steeper increases (or less negative declines). Researchers have argued that sociocultural changes, such as a greater emphasis on children's self-esteem by parents and teachers, grade inflation in educational systems, and increased possibilities for self-presentation through social media, have influenced the way in which children and adolescents perceive and evaluate themselves (Gentile et al., 2010; Twenge & Campbell, 2001, 2010). In Western countries such as the United States, there is empirical evidence for the above-mentioned sociocultural changes (Twenge & Campbell, 2010), so it is possible that these changes have influenced the development of self-views in children and adolescents. In addition to these substantive considerations, testing for cohort differences is also

important for methodological reasons. If mean-level change in self-evaluations differs across birth cohorts (with age held constant), then conclusions about normative development cannot be generalized across generations. Nevertheless, we note that a number of studies did not support the hypothesis that there have been generational changes in the positivity of self-perceptions (Trzesniewski & Donnellan, 2009, 2010; Trzesniewski, Donnellan, & Robins, 2008). Moreover, with regard to global self-esteem, cohort-sequential longitudinal studies suggest that the normative trajectory did not significantly change across the generations born in the 20th century (Erol & Orth, 2011; Orth, Robins, & Widaman, 2012; Orth et al., 2010; but see Twenge, Carter, & Campbell, 2017).

Trajectories of domain-specific self-evaluations could also differ by country and ethnicity. For example, theory suggests that individuals from Asian and Western cultures develop different self-construal styles, which may influence the typical trajectories of self-evaluations in these cultural contexts (Heine, Lehman, Markus, & Kitayama, 1999; Markus & Kitayama, 1991). More generally, researchers have pointed to the need to assess the degree to which psychological phenomena and processes hold across, or are unique to, different cultural contexts (Arnett, 2008; Henrich, Heine, & Norenzayan, 2010). Although the evidence on self-evaluations is based mostly on Western and White/European samples (Gentile et al., 2009; Wilgenbusch & Merrell, 1999), some studies have examined change in self-evaluations in Non-Western countries (X. Chen, He, & Li, 2004; King & McInerney, 2014; Liu, Wang, & Parkins, 2005; Wu, Watkins, & Hattie, 2010). Moreover, some studies with U.S. samples have tested whether change in self-evaluations differs by ethnicity (Brown et al., 1998) or have tracked the development of self-evaluations in Non-White samples (Diemer, Marchand, McKellar, &

Malanchuk, 2016; Harris et al., 2018). Therefore, we coded studies also with regard to country and ethnicity.

The Present Research

The goal of this research was to synthesize the available longitudinal data on mean-level change in domain-specific self-evaluations. We searched for studies with samples from all ages across the life span (i.e., from early childhood to old age). However, the coding of studies showed that the age in eligible studies was restricted to 5 to 28 years (see below for further information). In the analyses, we first examined the average trajectory of domain-specific self-evaluations, and then tested for moderators of the trajectory, including gender, birth year, country, and ethnicity.

Table 1 shows the eight domains of self-evaluations that were included in this meta-analysis. These domains were selected for the following reasons. First, we included those six domains that are central in theories on self-concept and domain-specific self-evaluations (e.g., Harter, 2012a; Marsh, 1990a; Shavelson, Hubner, & Stanton, 1976) and in key measures, such as Harter's Self-Perception Profiles (e.g., Harter, 2012c) and Marsh's Self-Description Questionnaires (e.g., Marsh, 1990b). Specifically, we included the domains of academic abilities, athletic abilities, physical appearance, social acceptance, morality (or behavioral conduct), and romantic relationships. Second, we included two additional domains, that is, mathematics and verbal abilities. Although these categories are considered subdomains of general self-evaluations in the academic domain (Shavelson et al., 1976), they are explicitly distinguished in some measures such as Marsh's Self-Description Questionnaires. Moreover, many primary studies have focused on mathematics and verbal abilities (e.g., King & McInerney, 2014; Musu-Gillette, Wigfield, Harring, & Eccles, 2015; Nagy et al., 2010). Table 1 provides information about

domain content and synonyms, as well as the corresponding subscales in the measures by Harter and Marsh. Moreover, the first column of Table 1 shows brief terms for the eight domains. We note that the focus of the present research is on trait, not state, domain-specific self-evaluations (for a measure of state self-evaluations, see Heatherton & Polivy, 1991).

The present meta-analysis advances the field by yielding robust insights into the normative trajectories of self-evaluations in important domains (no prior meta-analysis or systematic review is available on the topic). Also, the results will indicate whether domain-specific self-evaluations follow developmental trajectories that are similar to, or different from, the trajectory of global self-esteem (as determined in a recent meta-analysis; see Orth et al., 2018). Moreover, the moderator analyses will provide information about the robustness of the findings and, potentially, about factors that explain heterogeneity in the trajectories of domain-specific self-evaluations.

Method

The present meta-analysis used anonymized data and therefore was exempt from approval by the Ethics Committee of the authors' institution (Faculty of Human Sciences, University of Bern), in accordance with national law. Data, materials, and code are available on the Open Science Framework (https://osf.io/hnrv5/).

For the present meta-analysis, we used the same general procedures as in the meta-analysis on the development of global self-esteem reported in Orth et al. (2018). For reasons of clarity and completeness, we provide all relevant information on the present meta-analysis in this Method section, even if some of the information has already been described in Orth et al. (2018).

Table 1

Domains of Self-Evaluations and Corresponding Subscales in the Measures by Harter and Marsh

Domain	Content and synonyms	Subscales in measures by Harter	Subscales in measures by Marsh
Academic	Academic abilities, scholastic competence, intellectual abilities	PS: Cognitive Competence SPPC: Scholastic Competence SPPA: Scholastic Competence SPPCS: Scholastic Competence SPP-Adults: Intelligence	SDQ-I: General School SDQ-II: General School SDQ-III: Academic
Appearance	Physical appearance, physical attractiveness, body satisfaction, body esteem	PS: — SPPC: Physical Appearance SPPA: Physical Appearance SPPCS: Physical Appearance SPP-Adults: Physical Appearance	SDQ-I: Physical Appearance SDQ-II: Physical Appearance SDQ-III: Physical Appearance
Athletic	Athletic abilities, sports competences	PS: Physical Competence SPPC: Athletic Competence SPPA: Athletic Competence SPPCS: Athletic Competence SPP-Adults: Athletic Abilities	SDQ-I: Physical Abilities SDQ-II: Physical Abilities SDQ-III: Physical Abilities/Sports
Morality	Morality, honesty, behavioral conduct	PS: — SPPC: Behavioral Conduct SPPA: Behavioral Conduct SPPCS: Morality SPP-Adults: Morality	SDQ-I: — SDQ-II: Honesty-Trustworthiness SDQ-III: Honesty-Trustworthiness
Romantic	Romantic relationships, romantic appeal, opposite-sex relationships, intimate relationships	PS: — SPPC: — SPPA: Romantic Appeal SPPCS: Romantic Relationships SPP-Adults: Intimate Relationships	SDQ-I: — SDQ-II: Opposite-Sex Relations SDQ-III: Opposite-Sex Relations
Social	Social acceptance, social competence, sociability, popularity	PS: Peer Acceptance SPPC: Social Competence SPPA: Social Competence SPPCS: Social Acceptance SPP-Adults: Sociability	SDQ-I: Peer Relations SDQ-II: Same-Sex Relations SDQ-III: Same-Sex Relations

Mathematics	Mathematics abilities	PS: — SPPC: — SPPA: — SPPCS: — SPP-Adults: —	SDQ-I: Mathematics SDQ-II: Mathematics SDQ-III: Mathematics
Verbal	Verbal abilities, language, reading, literacy	PS: — SPPC: — SPPA: — SPPCS: — SPP-Adults: —	SDQ-I: Reading SDQ-II: Verbal SDQ-III: Verbal

Note. Information on Harter's and Marsh's measures is provided in the following sources: PS = Pictorial Scale of Perceived Competence and Social Acceptance for Young Children (Harter & Pike, 1984); SPPC = Self-Perception Profile for Children (Harter, 1982); SPPA = Self-Perception Profile for Adolescents (Harter, 2012b); SPPCS = Self-Perception Profile for College Students (Neemann & Harter, 2012); SPP-Adults = Self-Perception Profile for Adults (Messer & Harter, 2012); SDQ-I = Self-Description Questionnaire I (Marsh, Barnes, Cairns, & Tidman, 1984); SDQ-II = Self-Description Questionnaire II (Marsh, Ellis, Parada, Richards, & Heubeck, 2005); SDQ-III = Self-Description Questionnaire III (Marsh & O'Neill, 1984).

Selection of Studies

To search for relevant studies, we used three strategies. First, English-language journal articles, books, book chapters, and dissertations were searched in the database PsycINFO. A first search was conducted in September 2015 together with a search for articles on global self-esteem (Orth et al., 2018), which is the reason why we did not restrict the search to domain-specific measures of self-evaluations. We complemented the set of potentially relevant articles with a second search in October 2018 (i.e., by identifying articles included in PsycINFO after September 2015). Thus, the search of studies covered all entries in PsycINFO from 1806 to October 2018. We used the following search terms: *self-esteem, self-worth, self-concept, self-liking, self-respect, self-regard, self-acceptance, self-view*, and self-image**. The asterisk (i.e., the truncation symbol) allowed for the inclusion of alternate word endings of the search term (e.g., *self-view** yielded entries containing the term "self-view" but also "self-views"). We restricted the search to empirical-quantitative and longitudinal studies, by using the limitation options "empirical study," "quantitative study," and "longitudinal study" in PsycINFO. This search yielded 2,156 articles. Second, we examined the references cited in four narrative reviews of research on self-esteem development (Orth, 2017; Orth & Robins, 2014; Robins & Trzesniewski, 2005; Trzesniewski, Donnellan, & Robins, 2013) and cited in three meta-analyses using longitudinal data on self-esteem (Huang, 2010; Sowislo & Orth, 2013; Trzesniewski, Donnellan, & Robins, 2003). This search resulted in 77 additional potentially relevant articles. Third, we included all other relevant articles that we were aware of, resulting in 3 additional articles. Thus, overall, there were 2,236 potentially relevant articles.

To decide on the eligibility of studies, all articles were assessed in full text by the second, third, fourth, or fifth author.[1] In addition, 120 studies were rated by two of the authors to obtain

estimates of interrater agreement. The interrater agreement on inclusion or exclusion in the meta-analysis was high ($\kappa \geq .95$) and all diverging assessments were discussed until consensus was reached.

We included dissertations in the meta-analysis because dissertations are a category of the "gray" literature, providing a promising way to examine publication bias (Ferguson & Brannick, 2012; B. D. McLeod & Weisz, 2004). Although dissertations are publicly available and indexed in databases, Ferguson and Brannick (2012) argue that publication bias is less of an issue in dissertations because dissertations are typically submitted to dissertation committees regardless of whether the findings are statistically significant or not. Consistent with this reasoning, Ferguson and Brannick (2012) reported that effect sizes from dissertations typically yield effect sizes that differ more strongly from effect sizes from peer-reviewed journal articles than do effect sizes from unpublished manuscripts.

For inclusion in the present meta-analysis, we used the same set of criteria as in the meta-analysis by Orth et al. (2018): (a) the study used an explicit measure of domain-specific self-evaluations; (b) the study used a longitudinal study design (i.e., it included two or more assessments of the same sample); (c) the time lag between the first and last assessment was 6 months or longer (note that if a study included more than two assessments, each interval coded was at least 6 months or longer and intervals coded did not overlap); (d) the measure was identical across assessments (i.e., with regard to number of items, item wording, response scale, etc.); (e) the sample included at least 30 participants; (f) the sample was not a clinical sample; (g) the sample as a whole did not undergo a psychological or psychopharmacological intervention (i.e., the sample was not a treatment group of an intervention study; however, we used information from control groups if the control group did not undergo any alternative treatment);

and (h) enough information was given to compute effect sizes. Moreover, studies were included only if (i) the sample was sufficiently homogeneous with regard to age, as operationalized by a cutoff value of *SD* = 5 years for age at Time 1. This inclusion criterion was needed to ensure that the study can provide a valid estimate of age-related change in domain-specific self-evaluations. These procedures left 103 articles for analysis, providing effect sizes on 143 independent samples.

Coding of Studies

We coded the following data: year of publication, publication type, sample size, sample type, proportion of female participants, country in which sample was collected, ethnicity, year of Time 1 assessment, measure of domain-specific self-evaluations, mean age of participants at Time 1, standard deviation of age at Time 1, time lag between assessments, and effect size information.

If studies provided information that allowed coding of independent subsamples (e.g., female and male participants), we coded subsamples rather than the full sample to increase the precision of moderator analyses. If year of Time 1 assessment was not reported in the article or in other publications or sources of information on the sample, we estimated it using the following formula: Year of Time 1 assessment = publication year − 3 years (assuming that studies were published on average 3 years after the completion of data collection) − interval between first and last assessment (i.e., duration of data collection). If studies did not report the mean age of participants but valid indicators of age were given, we used this information to estimate age. For example, if a study reported that participants were children in 5th grade, we estimated mean age of participants as 11 years (thus, the general rule was adding the value of 6 to the grade). As reported above, studies were excluded if the standard deviation of Time 1 age was greater than 5

years (i.e., if the age variability in the sample was too large). Some studies did not report the standard deviation of age, although all other information needed for including the study was available. We included these studies if other information clearly suggested that the sample was sufficiently homogeneous with regard to age (e.g., if all participants were children in the same grade).

For studies that included more than two assessments, we coded all available assessments if the intervals between assessments were 6 months or longer. Later, in the meta-analytic computations, we ensured that each study provided only one effect size estimate per analysis. Thus, when a study provided more than one effect size for a given age period to be meta-analyzed (e.g., age 10 to 12 years), we first averaged effect sizes within studies (by computing the mean) and then conducted the meta-analytic computations. If a study included more than two assessments, but the intervals between assessments were shorter than 6 months, we used those assessments that provided for consecutive (i.e., non-overlapping) intervals that were at least 6 months long. For example, if a study included 5 assessments with 3-month-intervals, we used the first, third, and fifth assessment to compute effect sizes that were based on 6-month-intervals.

As the effect size measure, we used the standardized mean change *d* per year, denoted as d_{year} (Orth et al., 2018). We first computed the standardized mean change by subtracting the Time 1 mean from the Time 2 mean and dividing this difference by the Time 1 standard deviation (Morris & DeShon, 2002). Thus, computing standardized mean changes yielded *d* values (Cohen, 1988), with positive *d* values indicating an increase in domain-specific self-evaluations and negative *d* values indicating a decrease. Next, we set the standardized mean change in relation to the observed time interval, by dividing it by the length of the time lag between Time 1 and Time 2. Thus, the effect size measure used in the present meta-analysis is a

change-to-time ratio, with the unit *d* per year. If information on the means and standard deviations of domain-specific self-evaluations was not given in the article, but *d* values of mean-level change were reported, we used these to compute d_{year}.

The articles were coded by the second, third, fourth, or fifth author. In addition, 65 studies were coded by two of the authors to obtain estimates of interrater agreement. The interrater agreement was high ($\kappa \geq .97$ for categorical variables and $r \geq .97$ for continuous variables).[2] All diverging assessments were discussed until consensus was reached.

Meta-Analytic Procedure

The meta-analytic computations were made with R (R Core Team, 2019), using the metafor package (Viechtbauer, 2010). In the effect size analyses, we used random-effects models (for estimating weighted mean effect sizes) and mixed-effects meta-regression models (for testing moderators), following recommendations by Borenstein, Hedges, Higgins, and Rothstein (2009) and Raudenbush (2009). Between-study heterogeneity (i.e., τ^2) was estimated with restricted maximum likelihood estimation, as recommended by Viechtbauer (2005, 2010). Following Borenstein et al. (2009), study weights are given by

$$\omega_i = \left(\frac{1}{v_i + \tau^2}\right),$$

where ω_i is the study weight for study *i*, v_i is the within-study variance for study *i*, and τ^2 is the estimate of between-study heterogeneity. When using standardized mean change as effect size, the within-study variance is given by

$$v_i = \frac{2(1 - r_i)}{n_i} + \frac{{d_i}^2}{2n_i},$$

where d_i is the effect size in study i, n_i is the sample size in study i, and r_i is the correlation between pre- and post-scores in study i (Borenstein et al., 2009). Because this correlation is frequently not reported in primary studies, as in Orth et al. (2018) we used an estimate of .50, based on findings from a meta-analysis on test-retest stability in measures of self-esteem (Trzesniewski et al., 2003).

Results

Description of Studies

The meta-analytic dataset included 143 samples (Table 2 shows basic sample characteristics). Data were drawn from 98 journal articles and 5 dissertations. These 103 articles were published between 1984 and 2018, with the median in 2009. Sample sizes ranged from 36 to 20,644 (*M* = 784.6, *SD* = 1,915.4, *Mdn* = 317.0). In sum, the samples included 112,204 participants. Ninety-two percent of the samples were community samples, 6% were nationally representative, and 2% were samples of college students. The mean proportion of female participants was 54% (range = 0% to 100%, *SD* = 33%, *Mdn* = 51%). Forty percent of the samples were from the United States, 13% from Germany, 9% from Australia, 8% from China, 5% from Canada, 4% from Sweden, 4% from the Netherlands, 3% from Portugal, 3% from Switzerland, 2% from Norway, and the remaining 9% from other countries. Taken together, most samples (90%) were from Western cultural contexts such as the United States, European countries, Australia, and Canada; the remaining samples (10%) were from East and Southeast Asian countries including China, Indonesia, and Singapore; no African, South American, or Central American samples were included. With regard to ethnicity, 66% of the samples were

predominantly White/European ("predominantly" was defined as 80% and more), 11% predominantly Asian, 2% predominantly Black, 2% predominantly Hispanic/Latin American, 2% predominantly Native American, and 18% were other/mixed; the numbers do not add up to 100% because of rounding. Mean age at Time 1 ranged from 4.9 to 26.0 years (*M* = 12.0, *SD* = 3.5; note that some studies included 3 or more waves of data, and that we used the mean age at the center of time intervals for the effect size analyses, so the highest mean age examined was 27.5 years).[3] Year of Time 1 assessment ranged from 1966 to 2013 (*M* = 1998.3, *SD* = 8.9). We computed mean year of birth using the variables mean age at Time 1 and year of Time 1 assessment. Mean year of birth ranged from 1950 to 2007 (*M* = 1986.3, *SD* = 9.6). To assess domain-specific self-evaluations, 32% of the studies used one of the scales developed by Harter, 18% one of the scales by Marsh, and 50% another measure (for the scales developed by Harter and Marsh, see Table 1). The other measures included a broad range of measures. Among these, the most frequently used measure (with 7% of the samples) was the Body Esteem Scale for Adolescents and Adults (Mendelson, Mendelson, & White, 2001); all other measures were used in only a few (i.e., 3% or less) of the samples.

As reported in the Method section, one inclusion criterion for studies was that the sample was sufficiently homogeneous with regard to the age of participants (using a cutoff value of *SD* = 5 years for age at Time 1). This criterion was needed to ensure that effect sizes can be mapped with sufficient precision on age. Across samples, the standard deviation of age was relatively small with a mean of 0.74 years, ranging from 0.07 to 2.10 years. Thus, the findings suggest that age variability in the samples was not a concern in this meta-analysis.

Table 2
Descriptive Information on the Studies Included in the Meta-Analysis

Study	Sample size	Mean age at Time 1	*SD* of age at Time 1	Year of Time 1	Female	Sample type	Country	Ethnicity	Measure
Antonishak (2005)	169	13.36	0.66	1998	.53	Community	USA	Other	Harter
Antunes & Fontaine (2007), Cohort A	187	13.50	acc.	2001	1.00	Community	Portugal	White	Marsh
Antunes & Fontaine (2007), Cohort B	139	13.50	acc.	2001	.00	Community	Portugal	White	Marsh
Antunes & Fontaine (2007), Cohort C	167	15.50	acc.	2001	1.00	Community	Portugal	White	Marsh
Antunes & Fontaine (2007), Cohort D	123	15.50	acc.	2001	.00	Community	Portugal	White	Marsh
Arens et al. (2016), female subsample	205	4.87	0.36	2008	1.00	Community	Germany	White	Marsh
Arens et al. (2016), male subsample	215	4.87	0.36	2008	.00	Community	Germany	White	Marsh
Asendorpf & van Aken (1994)	50	9.00	acc.	1984	.47	Community	Germany	White	Harter
Bao & Jin (2015), control group	80	14.65	0.64	2009	.54	Community	China	Asian	Other
Becker & Neumann (2016)	155	10.00	acc.	2003	.52	Community	Germany	White	Other
Becker & Neumann (2018)	1,617	12.30	0.55	2011	.49	Community	Germany	White	Marsh
Becker et al. (2014), regular students	3,169	10.00	acc.	2003		Community	Germany	White	Other
Bellmore & Cillessen (2006)	491	12.38	0.67	2001	.49	Community	USA	Other	Other
Bodkin-Andrews et al. (2012)	1,211	13.50	acc.	2008	.49	Community	Australia	Other	Marsh
Bornholt & Piccolo (2005), Study 2	56	8.00	2.10	2001	.43	Community	Australia		Other
Bosacki (2015)	91	6.17	acc.	2003	.57	Community	Canada	White	Harter
Brown et al. (1998), Black subsample	472	9.00	acc.	1987	1.00	Community	USA	Black	Harter
Brown et al. (1998), White subsample	560	9.00	acc.	1987	1.00	Community	USA	White	Harter
Cairns et al. (1990)	2,543	17.00	acc.	1984	.53	Community	Ireland		Harter
Cantin & Boivin (2004)	142	12.50	0.43	1999	.53	Community	Canada		Harter
Cast & Cadwell (2007), female subsample	201	26.00	acc.	1991	1.00	Community	USA	White	Other
Cast & Cadwell (2007), male subsample	201	24.00	acc.	1991	.00	Community	USA	White	Other
Chapman & Tunmer (1997)	118	5.11	0.07	1993		Community	New Zeal.		Other
Chen & Jackson (2009), female subsample, age 12-13	79	12.50	acc.	2005	1.00	Community	China	Asian	Other
Chen & Jackson (2009), female subsample, age 14-15	83	14.50	acc.	2005	1.00	Community	China	Asian	Other
Chen & Jackson (2009), female subsample, age 16-17	158	16.50	acc.	2005	1.00	Community	China	Asian	Other
Chen & Jackson (2009), male subsample, age 12-13	43	12.50	acc.	2005	.00	Community	China	Asian	Other
Chen & Jackson (2009), male subsample, age 14-15	61	14.50	acc.	2005	.00	Community	China	Asian	Other
Chen & Jackson (2009), male subsample, age 16-17	77	16.50	acc.	2005	.00	Community	China	Asian	Other
Chen et al. (2004), female subsample	258	12.40	0.67	1998	1.00	Community	China	Asian	Harter
Chen et al. (2004), male subsample	248	12.40	0.67	1998	.00	Community	China	Asian	Harter
Clark & Tiggemann (2008)	150	10.28	1.04	2003	1.00	Community	Australia	White	Other
Crocker et al. (2006)	501	14.50	acc.	1998	1.00	Community	Canada		Other
Dapp & Roebers (2018), female subsample	71	6.50	0.35	2013	1.00	Community	Switzerland	White	Other
Dapp & Roebers (2018), male subsample	84	6.50	0.35	2013	.00	Community	Switzerland	White	Other

Davis-Kean et al. (2008), Childhood and Beyond, Cohort 1	317	6.00	acc.	1987	.50	Community	USA	White	Other
Davis-Kean et al. (2008), Childhood and Beyond, Cohort 2	330	7.00	acc.	1987	.50	Community	USA	White	Other
Davis-Kean et al. (2008), Childhood and Beyond, Cohort 3	423	9.00	acc.	1987	.50	Community	USA	White	Other
Davison et al. (2007)	178	11.38	0.28	2001	1.00	Community	USA	White	Other
Davison et al. (2008)	163	9.34	0.30	2000	1.00	Community	USA	White	Harter
De Neve & Oswald (2012)	20,644	15.00	1.77	1994	.49	National	USA	Other	Other
Denny (2011)	6,966	9.00	acc.	2001	.52	National	USA	Other	Marsh
Dickhäuser et al. (2017)	1,641	10.50	0.43	2012	.53	Community	Germany	White	Other
Diemer et al. (2016)	618	14.00	acc.	1993	.46	Community	USA	Black	Other
Dohnt & Tiggemann (2006)	97	6.91	1.23	2002	1.00	Community	Australia	White	Harter
Donnellan et al. (2007)	409	23.26	0.47	1999	.58	Community	USA	White	Harter
Dufner et al. (2015)	709	11.83	1.99	2009	.54	Community	Germany	White	Harter
Eisenberg et al. (2006), female subsample, high school	440	12.70	0.74	1998	1.00	Community	USA	Other	Other
Eisenberg et al. (2006), female subsample, young adults	946	15.80	0.81	1998	1.00	Community	USA	Other	Other
Eisenberg et al. (2006), male subsample, high school	366	12.80	0.76	1998	.00	Community	USA	Other	Other
Eisenberg et al. (2006), male subsample, young adults	764	15.90	0.78	1998	.00	Community	USA	Other	Other
Fenzel (2000)	116	10.80	acc.	1996	.56	Community	USA	White	Harter
Ferreiro et al. (2012), female subsample	465	10.84	0.74	2005	1.00	Community	Spain	White	Other
Ferreiro et al. (2012), male subsample	477	10.83	0.75	2005	.00	Community	Spain	White	Other
Frisen et al. (2015), female subsample	515	10.00	acc.	2000	1.00	Community	Sweden	White	Other
Frisen et al. (2015), male subsample	445	10.00	acc.	2000	.00	Community	Sweden	White	Other
Gest et al. (2005)	400	10.00	acc.	2001	.56	Community	USA	White	Harter
Gestsdottir et al. (2016)	385	15.00	acc.	2005	.49	National	Iceland	White	Other
Gniewosz et al. (2012)	1,953	10.90	0.63	1983	.53	Community	USA	White	Other
Gruenenfelder-Steiger et al. (2016)	1,023	12.00	acc.	1979	.48	Community	Germany	White	Other
Guay et al. (1999)	397	9.00	acc.	1995	.52	Community	Canada		Harter
Guo et al. (2015)	2,213	16.00	acc.	1966	.00	National	USA	Other	Other
Harris et al. (2018)	674	10.80	0.61	2006	.50	Community	USA	Hispanic	Marsh
Helmke & van Aken (1995)	697	8.00	acc.	1990	.49	Community	Germany	White	Other
Hoge et al. (1990)	322	12.00	acc.	1983	.55	Community	USA	White	Other
Hoglund (1995)	39	10.00	acc.	1987	.53	Community	USA	White	Other
Impett et al. (2011), Study 1	183	14.00	acc.	1998	1.00	Community	USA	Other	Other
Impett et al. (2011), Study 2	133	14.00	acc.	2001	1.00	Community	USA	Other	Other
Ireson & Hallam (2009)	1,687	13.50	acc.	2004		Community	UK	White	Marsh
Keel et al. (1997), female subsample	80	11.50	acc.	1993	1.00	Community	USA	White	Other
Keel et al. (1997), male subsample	85	11.50	acc.	1993	.00	Community	USA	White	Other
Kimber et al. (2008), control group, senior sample	61	13.50	acc.	2001		Community	Sweden	White	Other
King & McInerney (2014), female subsample	1,178	12.23	0.65	2010	1.00	Community	China	Asian	Marsh
King & McInerney (2014), male subsample	1,440	12.23	0.65	2010	.00	Community	China	Asian	Marsh
Kuzucu et al. (2014)	406	9.00	acc.	1975	.46	Community	USA	Other	Harter

LaGrange et al. (2008), Cohort 1	187	8.45	0.65	2003	.56	Community	USA	Other	Harter
LaGrange et al. (2008), Cohort 2	169	9.56	0.67	2003	.56	Community	USA	Other	Harter
LaGrange et al. (2008), Cohort 3	171	11.35	0.56	2003	.56	Community	USA	Other	Harter
Lamote et al. (2014)	3,900	14.00	acc.	1991	.43	Community	Belgium	White	Other
Le Bars et al. (2009), Study 2	82	16.70	1.10	2004	.45	Community	France	White	Other
Liu et al. (2005)	495	13.00	acc.	1999	.52	Community	Singapore	Asian	Other
Lüdtke et al. (2005)	2,141	13.70	acc.	1995	.50	National	Germany	White	Other
Luszczynska & Abraham (2012)	551	16.43	0.60	2008	.58	Community	Poland	White	Other
Mantzicopoulos (2006), preschool and kindergarten sample	87	5.50	0.32	2002	.52	Community	USA	Other	Harter
Mantzicopoulos (2006), first and second grade sample	87	7.50	0.32	2002	.52	Community	USA	Other	Harter
Marsh et al. (1998), Kindergarten	127	5.40	0.40	1994		Community	Australia		Marsh
Marsh et al. (1998), Grade 1	139	6.30	0.40	1994		Community	Australia		Marsh
Marsh et al. (1998), Grade 2	130	7.40	0.50	1994		Community	Australia		Marsh
Marsh et al. (2005), Study 3	3,731	15.00	acc.	2000		Community	Australia		Marsh
McGrath & Repetti (2002)	246	9.50	acc.	1997	.47	Community	USA	White	Harter
McGuire et al. (1999)	496	12.90	2.00	1989	.51	Community	USA	Other	Harter
McKinley (2006), female subsample	115	18.97	1.37	1993	1.00	College	USA	White	Other
McKinley (2006), male subsample	49	19.40	1.81	1993	.00	College	USA	White	Other
McLeod & Owens (2004)	547	10.50	acc.	1986	.51	National	USA	Other	Harter
Modecki et al. (2018)	1,146	13.29	acc.	2011	.55	Community	Australia	White	Marsh
Möller et al. (2011)	1,508	11.05	0.56	2006	.49	Community	Germany	White	Marsh
Möller et al. (2014)	1,045	11.00	acc.	2004	.50	Community	Germany	White	Other
Morgan et al. (2012), control group	50	14.20	0.40	2009		Community	Australia	White	Other
Morin et al. (2011)	1,001	12.62	0.63	2000	.46	Community	Canada	White	Marsh
Mössle & Rehbein (2013)	668	11.50	acc.	2008	.49	Community	Germany	White	Other
Mullins (1997), female subsample	131	11.00	acc.	1992	1.00	Community	USA	White	Harter
Mullins (1997), male subsample	120	11.00	acc.	1992	.00	Community	USA	White	Harter
Musu-Gilette et al. (2015)	421	10.00	acc.	1987		Community	USA	Other	Other
Nagy et al. (2010), Australian subsample	744	13.00	acc.	1995	.44	Community	Australia	Other	Marsh
Nagy et al. (2010), German subsample	3,533	13.00	acc.	1991	.58	Community	Germany	White	Other
Nagy et al. (2010), US subsample	2,266	13.00	acc.	1983	.53	Community	USA	White	Other
Nelson et al. (2018)	967	10.36	0.52	2000	.53	Community	Sweden	White	Other
Newman (1984), female subsample	60	7.40	0.31	1973	1.00	Community	USA		Other
Newman (1984), male subsample	81	7.40	0.31	1973	.00	Community	USA		Other
Niepel et al. (2014), Study 1	1,529	10.69	0.44	2007	.47	Community	Germany	White	Marsh
Niepel et al. (2014), Study 2	639	10.70	0.43	2008	.44	Community	Germany	White	Marsh
Nogueira Avelar e Silva et al. (2018), female subsample	358	13.30	0.57	2011	1.00	Community	Netherlands	White	Harter
Nogueira Avelar e Silva et al. (2018), male subsample	358	13.30	0.58	2011	.00	Community	Netherlands	White	Harter
O'Dea (2006)	80	12.80	0.60	2001	1.00	Community	Australia	White	Harter
O'Dea & Abraham (2000), control group	195	12.90	0.60	1996	.63	Community	Australia		Harter

Ohannessian et al. (1996), female subsample	103	12.20	0.68	1990	1.00	Community	USA	White	Harter
Ohannessian et al. (1996), male subsample	101	12.20	0.68	1990	.00	Community	USA	White	Harter
Ohannessian et al. (2019)	636	16.10	0.71	2007	.54	Community	USA	Other	Harter
Orth et al. (2014)	672	10.40	0.60	2006	.50	Community	USA	Hispanic	Marsh
Pomerantz & Dong (2006)	126	10.03	0.83	1997	.44	Community	USA	White	Other
Preckel et al. (2013)	1,282	11.02	0.44	2007	.48	Community	Germany	White	Other
Radin (2005), younger cohort	290	11.67	0.69	1988	.50	Community	USA	Native	Harter
Radin (2005), older cohort	283	13.69	0.72	1990	.50	Community	USA	Native	Harter
Raustorp et al. (2009), female subsample	36	12.70	acc.	2000	1.00	Community	Sweden	White	Other
Raustorp et al. (2009), male subsample	41	12.70	acc.	2000	.00	Community	Sweden	White	Other
Roebers et al. (2012)	209	7.50	0.33	2008	.52	Community	Switzerland	White	Other
Sallquist et al. (2010)	205	13.47	0.69	2004	.55	Community	Indonesia	Asian	Harter
Schneider et al. (2008), control group	59	15.02	0.77	2004	1.00	Community	USA		Marsh
Shapka & Keating (2005)	518	15.50	acc.	2000	.49	Community	Canada	White	Harter
Siffert et al. (2012)	176	10.61	0.40	2008	.51	Community	Switzerland	White	Harter
Silverthorn et al. (2005)	342	14.00	acc.	1999	.50	Community	Canada		Other
Slutzky & Simpkins (2009)	987	9.55	1.31	1989	.51	Community	USA	White	Other
Spray et al. (2013)	491	11.29	0.30	2009	.51	Community	UK	White	Marsh
Steiger et al. (2014)	1,527	12.00	acc.	1979	.51	Community	Germany	White	Other
Udell et al. (2010), female subsample	258	13.31	0.51	2002	1.00	Community	Netherlands	White	Harter
Udell et al. (2010), male subsample	212	13.31	0.51	2002	.00	Community	Netherlands	White	Harter
Valkenburg et al. (2017)	852	12.50	1.36	2012	.51	Community	Netherlands	White	Harter
Viljaranta et al. (2014)	216	7.00	acc.	2000	.48	Community	Finland	White	Other
von Soest et al. (2016)	3,116	15.50	acc.	1992		National	Norway	White	Harter
Wichstrom & von Soest (2016), female subsample	1,769	15.10	2.05	1992	1.00	National	Norway	White	Harter
Wichstrom & von Soest (2016), male subsample	1,482	14.88	1.80	1992	.00	National	Norway	White	Harter
Wu et al. (2010)	1,044	15.00	1.70	2006	.48	Community	China	Asian	Other
Young & Mroczek (2003)	261	15.50	acc.	1999	.45	College	USA	White	Harter

Note. Mean age and standard deviation of age are given in years. As described in the Method section, some studies did not report the standard deviation of age, but other information clearly suggested that the sample was sufficiently homogeneous with regard to age (e.g., all participants were children in the same grade). For these studies, the standard deviation is denoted as acceptable (“acc.”). The column “Female” shows the proportion of female participants.

As also reported above, some studies provided effect sizes for more than one age. The reason is that some of the longitudinal studies included more than two waves of data, allowing computation of effect sizes for more than one interval. Specifically, the number of intervals ranged from 1 to 7 across studies. Because the goal of the meta-analysis was to comprehensively summarize all available data on mean-level change in domain-specific self-evaluations, it was important not to ignore information that multi-wave studies provided at later waves (i.e., Waves 3 and later). With regard to these multi-wave studies, the following two procedures should be noted. First, as described earlier, if a study included more than two assessments, each interval coded was at least 6 months or longer (as also required for 2-wave studies) and intervals coded from the same study did not overlap. Second, we ensured that all meta-analytic computations were conducted with independent samples (i.e., no participant provided information for more than one effect size included in the same analysis). Therefore, for each of the analyses, we first averaged effect sizes within studies and then conducted the meta-analytic computations. For this reason, we had to use different datasets depending on the specific analysis. In the moderator analyses, we used information from all 143 samples that provided effect sizes, by first averaging effect sizes within studies. In contrast, in the effect size analyses, which were conducted within age groups, we averaged effect sizes from multi-wave studies only within the specific age range.

Preliminary Analyses

We searched for potential outliers using the “influence” command of the metafor package (Viechtbauer, 2010). According to Viechtbauer and Cheung (2010), studies with absolute studentized deleted residuals larger than 1.96 should be inspected more closely. We examined whether there was evidence that these studies were influential by assessing their DFFITS values (i.e., difference between the predicted average effect for the study with vs. without including it in

model fitting; Viechtbauer & Cheung, 2010). There were very few relevant cases. For three of the domains of self-evaluations, there was no relevant case (mathematics, morality, romantic), for four domains there was one relevant case (academic, appearance, athletic, verbal), and for one domain there were two relevant cases (social). Given the relatively large number of effect sizes for most of the domains (see Table 3), the findings suggest that the influence of potential outliers was probably small. Moreover, we inspected the effect size codings of these studies, which did not suggest that there were any errors or implausible values in the effect sizes. We therefore retained all effect sizes in the meta-analytic dataset, which is consistent with methodological literature advising against routine deletion of studies with particularly large or small effect sizes (Viechtbauer & Cheung, 2010).

Next, we assessed whether there was evidence of publication bias in the data. We expected that publication bias would not be a problem in this meta-analysis because many studies included did not focus on mean-level change in domain-specific self-evaluations (i.e., they examined other research questions), but simply reported the relevant statistics (i.e., means and standard deviations of domain-specific self-evaluations) together with statistics on a larger set of variables. We used three methods to examine publication bias. First, for each of the domains of self-evaluations we examined the funnel graph, which displays the relation between effect size and the standard error of the effect size (Sutton, 2009). The funnel graphs exhibited a symmetrical shape typical of nonbiased meta-analytic datasets (Figure 1). Second, for each of the domains we conducted Egger's regression test (Egger, Smith, Schneider, & Minder, 1997). Since we conducted eight tests but did not expect publication bias in the present data (thus, 1 out of 20 tests would be expected to be significant by chance on the .05 level), we adjusted the significance level to $p < .0063$, following the Bonferroni method (i.e., dividing .05 by 8). The

results showed that for all domains Egger's test was nonsignificant, suggesting that the funnel graph did not deviate significantly from a symmetrical shape (Table 3). Third, we compared effect sizes from dissertations (as a category of gray literature) with effect sizes from peer-reviewed journal articles, by using mixed-effects meta-regression models. If dissertations yield effect sizes that differ significantly from journal articles, this is evidence for publication bias. For two of the domains of self-evaluations (athletic and romantic), no effect sizes from dissertations were available, so the test could not be conducted for these domains. For the remaining domains, the differences between journal articles and dissertations were nonsignificant (Table 3). For three domains (mathematics, morality, and verbal), only one or two effect sizes from dissertations were available; thus, the conclusions that can be drawn with regard to these domains are probably limited. Nevertheless, we note that all three methods used converged in suggesting that there was no evidence of publication bias.

Table 3

Tests of Publication Bias in Mean-Level Change (d_{year}) in Domain-Specific Self-Evaluations

	Egger's regression test			Peer-reviewed journal articles versus dissertations[a]			
Domain	k	z	p	k_j	k_d	z	p
Academic	75	0.802	.422	65	10	−0.398	.690
Appearance	108	−0.494	.621	102	6	−0.981	.327
Athletic	45	0.801	.423	45	0	—	—
Morality	20	−1.210	.226	18	2	0.173	.863
Romantic	11	−0.864	.388	11	0	—	—
Social	72	−1.302	.193	63	9	−0.817	.414
Mathematics	60	1.477	.140	59	1	−0.202	.840
Verbal	27	2.168	.030	26	1	−1.137	.255

Note. Regression coefficients are unstandardized. Dash indicates that model cannot be estimated due to lack of effect sizes from dissertations. d_{year} = standardized mean change d per year; k = number of effect sizes; k_j = number of effect sizes from peer-reviewed journal articles; k_d = number of effect sizes from dissertations.

[a] 1 = dissertation, 0 = peer-reviewed journal article.

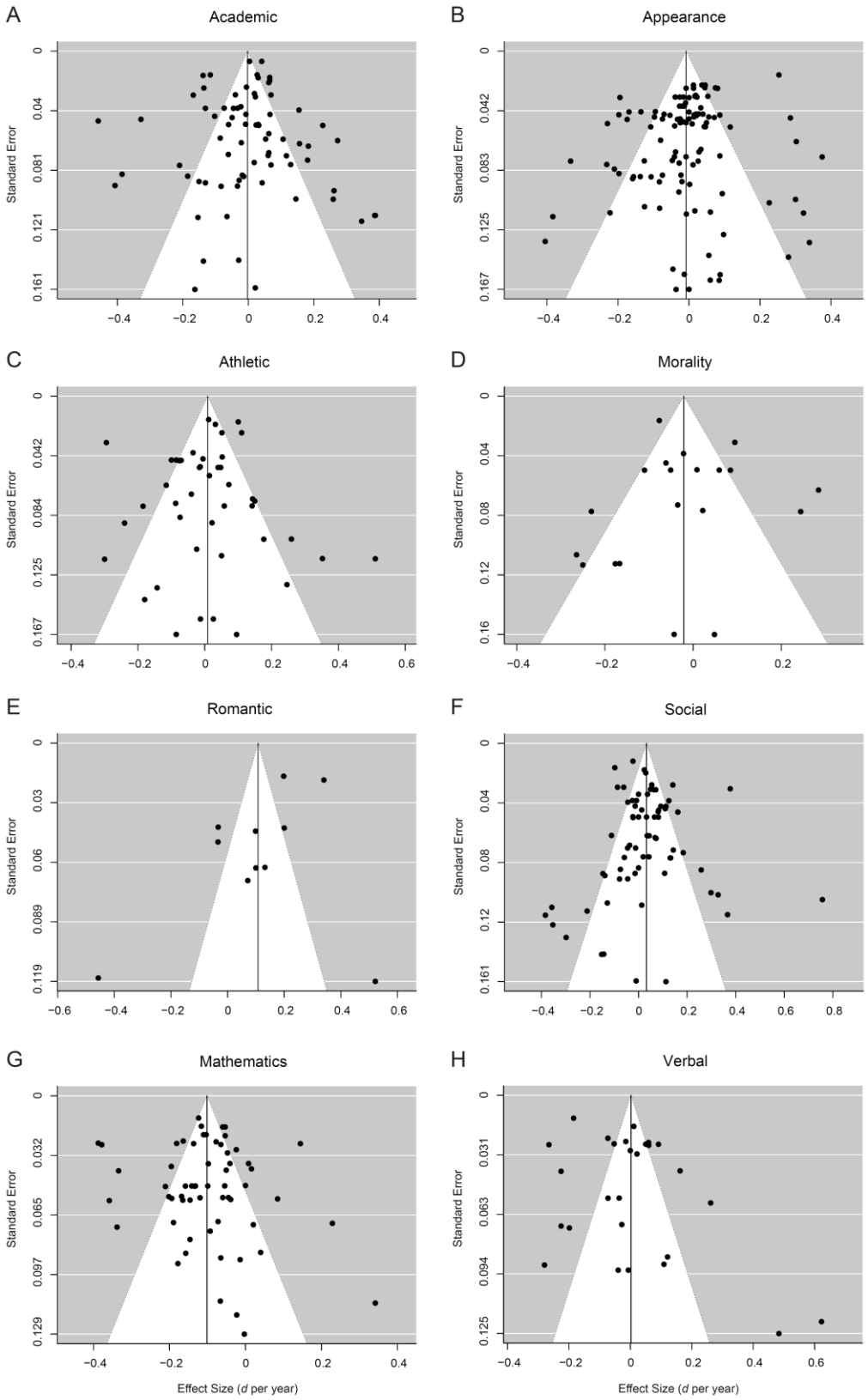

Figure 1. Funnel graphs displaying the relation between standard error and effect size.

Effect Size Analyses

As the goal of this meta-analysis was to map mean-level change in domain-specific self-evaluations on age, effect size analyses were conducted within age groups. For these analyses, we constructed multiple age groups across the observed age range (see Table 4). As in Orth et al. (2018), we did not use the mean age at Time 1 as age variable but, instead, the mean age at the center of the time interval on which the effect size was based. For example, if a sample was assessed at age 10 years at Time 1 and age 12 years at Time 2, the age at the center of the interval on which the effect size was based was 11 years. Although the difference between the two age variables (i.e., age at Time 1 and age at the center of the interval) may be irrelevant for short intervals (e.g., 1 year), the difference is more relevant for long intervals (e.g., 4 years). Because mean-level changes in domain-specific self-evaluations might change systematically across long intervals (e.g., the slope might become smaller or larger with age), the most meaningful age value related to the observed effect size is the center of the Time 1–Time 2 interval rather than age at the beginning (Time 1) or end (Time 2) of the interval.

For the age range from 6 to 18 years, we constructed 2-year age groups. Above age 18, few studies were available for most domains, so we constructed one age group from 18 to 22 years (i.e., the college years) and another age group from 22 to 28 years (i.e., all other years in emerging adulthood for which data were available). The youngest age group covered a 1-year period only (5 to 6 years); we did not combine effect sizes for age 5 to 6 years with the age group from 6 to 8 years, because in many countries and school systems, 6 years marks the transition from kindergarten (or no schooling) to school.

Table 4
Estimates of Mean-Level Change in Domain-Specific Self-Evaluations

Domain and age (years)	k	N	Weighted mean effect size (d_{year})	95% CI	Heterogeneity Q	τ^2	I^2
Academic							
5–6	1	127	−0.131	[−.306, .044]	—	—	—
6–8	3	356	0.068	[−.243, .378]	14.8*	0.066	88.2
8–10	5	1,127	0.148*	[.003, .292]	16.7*	0.020	79.6
10–12	16	6,695	−0.001	[−.070, .067]	56.1*	0.014	82.1
12–14	21	9,442	−0.031	[−.111, .048]	187.3*	0.030	92.7
14–16	15	14,630	−0.071*	[−.119, −.022]	75.2*	0.006	84.7
16–18	9	34,999	0.048*	[.015, .082]	24.5*	0.002	83.1
18–22	2	20,927	0.131	[−.132, .395]	19.5*	0.034	94.9
22–28	3	811	0.043	[−.034, .120]	2.2	0.001	17.3
Appearance							
5–6	1	127	−0.159	[−.334, .016]	—	—	—
6–8	4	457	0.049	[−.190, .289]	19.5*	0.051	84.8
8–10	4	1,618	0.008	[−.095, .111]	13.9*	0.008	76.3
10–12	17	6,408	−0.060	[−.125, .005]	113.9*	0.015	84.3
12–14	24	7,387	−0.028	[−.083, .026]	84.1*	0.013	78.4
14–16	27	15,352	0.011	[−.044, .066]	260.8*	0.016	89.5
16–18	15	6,612	0.017	[−.013, .047]	18.7	0.001	21.7
18–22	8	7,273	0.017	[−.006, .040]	1.9	0.000	0.0
22–28	8	5,193	−0.004	[−.031, .023]	1.7	0.000	0.0
Athletic							
5–6	1	127	0.022	[−.152, .196]	—	—	—
6–8	3	356	0.060	[−.381, .501]	28.1*	0.142	94.0
8–10	4	731	0.047	[−.038, .132]	4.4	0.001	17.5
10–12	5	2,219	−0.094	[−.230, .042]	45.6*	0.021	89.0
12–14	11	3,548	0.054	[−.009, .117]	26.2*	0.007	64.0
14–16	11	7,183	−0.014	[−.070, .042]	25.4*	0.004	66.8
16–18	9	8,046	0.028	[−.031, .087]	31.8*	0.005	78.0
18–22	0	—	—	—	—	—	—
22–28	1	409	−0.014	[−.110, .083]	—	—	—
Morality							
5–6	0	—	—	—	—	—	—
6–8	1	91	−0.264*	[−.474, −.055]	—	—	—
8–10	2	593	0.037	[−.078, .152]	1.8	0.003	45.3
10–12	5	1,457	−0.071	[−.158, .017]	8.4	0.005	55.0
12–14	4	696	0.019	[−.187, .224]	13.5*	0.034	81.8
14–16	7	6,098	−0.010	[−.124, .104]	52.9*	0.020	92.1
16–18	0	—	—	—	—	—	—
18–22	0	—	—	—	—	—	—
22–28	1	409	0.009	[−.088, .106]	—	—	—

Romantic							
5–6	0	—	—	—	—	—	—
6–8	0	—	—	—	—	—	—
8–10	0	—	—	—	—	—	—
10–12	0	—	—	—	—	—	—
12–14	4	1,116	0.203*	[.031, .376]	12.5*	0.025	85.6
14–16	4	4,638	−0.025	[−.296, .246]	54.2*	0.072	97.4
16–18	2	3,634	0.223	[−.013, .459]	25.5*	0.028	96.1
18–22	0	—	—	—	—	—	—
22–28	1	409	−0.034	[−.131, .063]	—	—	—
Social							
5–6	1	127	−0.138	[−.313, .036]	—	—	—
6–8	5	511	−0.147	[−.301, .007]	11.4*	0.020	66.7
8–10	6	1,759	0.052	[−.044, .148]	15.9*	0.009	70.7
10–12	15	11,793	0.093*	[.006, .180]	105.7*	0.024	92.7
12–14	23	9,130	0.066*	[.027, .105]	60.7*	0.005	65.6
14–16	14	10,165	−0.044	[−.092, .004]	57.7*	0.005	76.9
16–18	4	4,824	0.119	[−.062, .300]	107.9*	0.038	97.1
18–22	0	—	—	—	—	—	—
22–28	4	3,927	0.013	[−.019, .045]	1.7	0.000	0.0
Mathematics							
5–6	3	547	0.084	[−.067, .234]	6.4*	0.012	68.0
6–8	8	1,496	−0.038	[−.158, .082]	29.7*	0.023	80.6
8–10	6	1,807	−0.124*	[−.207, −.041]	13.4*	0.006	61.1
10–12	9	13,774	−0.171*	[−.286, −.056]	318.4*	0.029	97.2
12–14	12	11,092	−0.097*	[−.155, −.040]	65.8*	0.009	88.7
14–16	13	16,562	−0.099*	[−.128, −.069]	31.9*	0.001	64.3
16–18	8	8,019	−0.096*	[−.145, −.048]	18.1*	0.003	69.4
18–22	1	421	−0.134*	[−.230, −.038]	—	—	—
22–28	0	—	—	—	—	—	—
Verbal							
5–6	2	245	0.053	[−.072, .179]	0.8	0.000	0.0
6–8	7	967	0.088	[−.157, .333]	62.3*	0.100	92.9
8–10	1	216	−0.225*	[−.360, −.090]	—	—	—
10–12	5	12,595	−0.139*	[−.251, −.027]	103.1*	0.016	96.9
12–14	8	9,278	0.068*	[.007, .130]	44.6*	0.007	88.5
14–16	3	5,760	0.001	[−.025, .027]	1.3	0.000	0.0
16–18	1	342	−0.073	[−.179, .033]	—	—	—
18–22	0	—	—	—	—	—	—
22–28	0	—	—	—	—	—	—

Note. Computations were made with random-effects models. Dash indicates that no data are available (if $k = 0$) or that estimate is not applicable (if $k = 1$). k = number of samples; N = total number of participants in the k samples; d_{year} = standardized mean change d per year; CI = confidence interval; Q = statistic used in heterogeneity test; τ^2 = estimated amount of total heterogeneity; I^2 = ratio of total heterogeneity to total variability (given in percent).
* $p < .05$.

Although the power of significance tests of mean-level change would be greater if we constructed broader age groups (which would then include a larger number of samples), it is important to emphasize that null-hypothesis significance testing of mean-level change was not a central goal in this meta-analysis (cf. Cumming, 2014; Fraley & Marks, 2007; Greenwald, 1975). In the present research, the goal was rather to obtain estimates of age-dependent mean-level change and, thus, narrower age groups provide more precision with regard to age. We used the weighted mean effect size (i.e., the point estimate) as best estimate of mean-level change in the age group, regardless of whether the estimate differed significantly from zero or not. In Table 4, we report the null-hypothesis significance tests of mean effect sizes for reasons of completeness.

Table 4 reports the meta-analytic findings for all domains of self-evaluations. Figure 2 illustrates the findings by aggregating the point estimates of mean-level change across the observed age range for each of the domains. The vertical axes show cumulative *d* values of domain-specific self-evaluations. Note that the cumulative *d*s are relative to age 5, so the point of origin (i.e., zero) is arbitrary. For age groups that covered more than one year (e.g., age 6–8; age 18–22), the estimate of yearly change (i.e., d_{year}) was used for each year included in the age group. Moreover, for age groups for which no effect sizes were observed (e.g., age 18 to 22 years in the athletic domain), the graphs are interrupted and continued with the next age group for which information is available.

Figure 2 illustrates that the trajectories of self-evaluations differed substantially across domains. We first focus on domains in which self-evaluations generally showed positive developmental trajectories. In the academic domain (Figure 2A), self-evaluations generally increased, despite slight decreases from age 5 to 6 years, and again from age 12 to 16 years. The figure shows that the strongest increases in academic self-evaluations occurred from age 6 to 10

years (i.e., during roughly the first four years of school) and from age 18 to 22 years (i.e., during the years when many individuals attend college). The net increase from age 5 to 28 years corresponded to almost one standard deviation (cumulative $d = 0.97$). In the social domain (Figure 2F), self-evaluations also increased at most ages. However, from age 5 to 8 years there was a relatively strong decrease ($d = -0.43$). Then, self-evaluations recovered and increased by $d = 0.57$ from age 8 to 18 years, with a slight dip at ages 14 to 16 years. Due to the initial decrease from age 5 to 8, the net increase from age 5 to 28 years was relatively small (cumulative $d = 0.22$). The third domain for which the overall trend was positive was the romantic domain (Figure 2E). For this domain, the youngest age group for which data were available was 12 to 14 years. From age 12 to 18 years, self-evaluations increased by $d = 0.80$, corresponding to a large effect size according to the guidelines by Cohen (1988). Later, from age 22 to 28 years, self-evaluations slightly declined in this domain ($d = -0.20$).

Next, there were two domains in which self-evaluations showed little mean-level change across the observed age range. In the appearance domain (Figure 2B), self-evaluations showed some small ups and downs, but the effect sizes of the increases and decreases were relatively small. For example, although self-evaluations of physical appearance declined from age 10 to 14 years (i.e., in prepubertal and pubertal years), the average decline was benign ($d = -0.17$). The net change from age 5 to 28 years was very small in this domain (cumulative $d = -0.12$). Similarly, in the athletic domain (Figure 2C) there were only small mean-level changes between ages 5 and 28 years. Although self-evaluations increased slightly from age 5 to 10 years ($d = 0.24$), later decreases and increases were even smaller in this domain, with cumulative $d = 0.10$ at age 28 years.

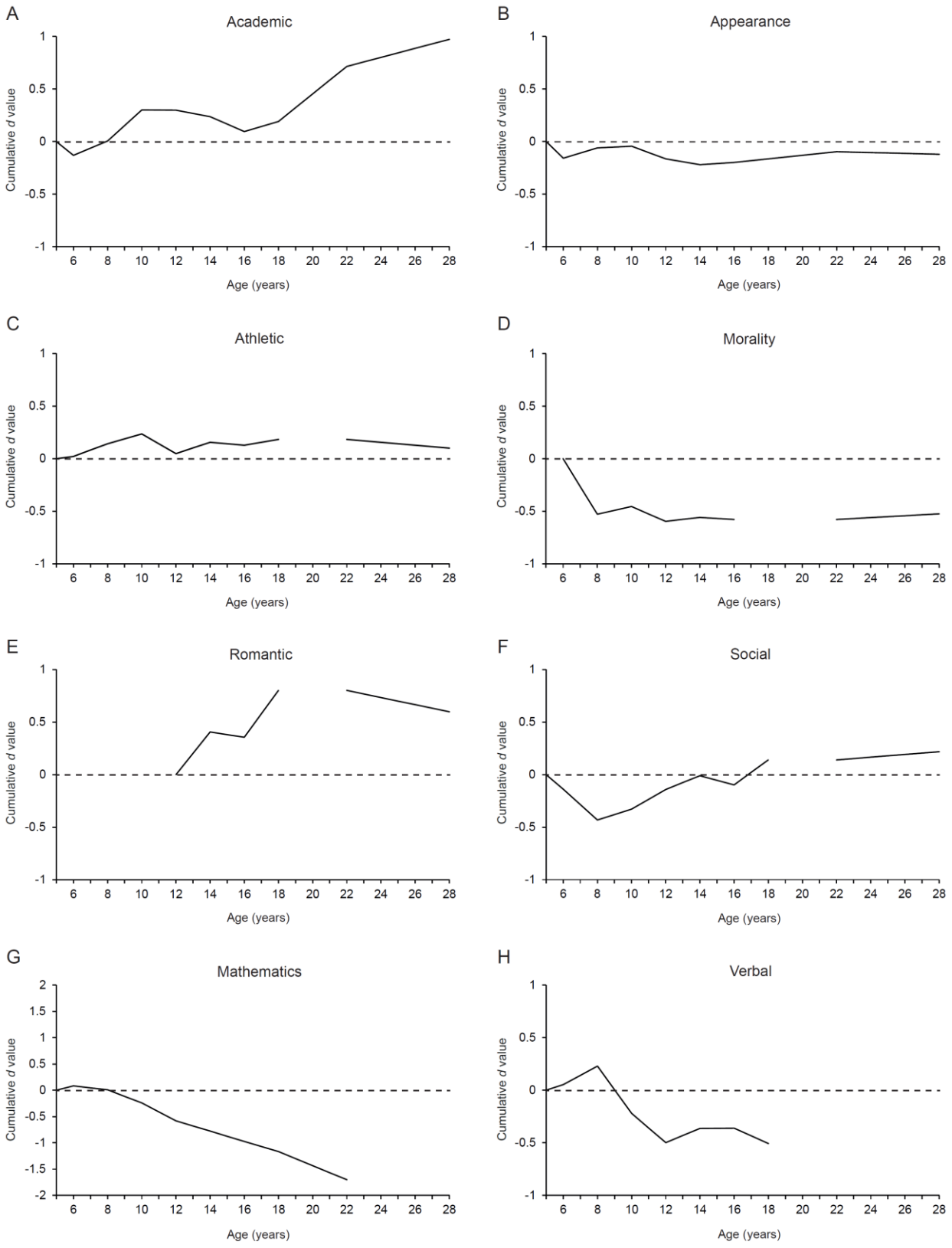
A
Academic
B
Appearance
C
Athletic
D
Morality
E
Romantic
F
Social
G
Mathematics
H
Verbal
Cumulative d value
Age (years)

Figure 2. Mean-level change in domain-specific self-evaluations from age 5 to 28 years. The figure shows cumulative *d* values relative to age 5 years; thus, the point of origin (i.e., zero) is arbitrary. For age groups for which no effect sizes were observed (e.g., age 18 to 22 years in Panel C), the graphs are interrupted and continued with the next age group for which information is available.

Finally, in some domains self-evaluations generally showed negative developmental trajectories. In the morality domain (Figure 2D), there was a relatively strong decline from age 6 to 8 years, corresponding to one half of a standard deviation over two years ($d = -0.53$). In later years, there were only small mean-level changes, but, if anything, self-evaluations did not show any important increases in this domain. The strongest decrease emerged in the domain of mathematics (Figure 2G). Except for a slight increase from age 5 to 6 years ($d = 0.08$), self-evaluations decreased in this domain, at a relatively stable rate from age 6 to 22 years. The net decrease from age 5 to 6 years corresponded to a very large effect size (cumulative $d = -1.70$; note that the figure uses a different scale on the vertical axis for this domain). In the domain of verbal abilities (Figure 2H), average levels of self-evaluations also showed an overall negative trend. However, in this domain, self-evaluations first increased from age 5 to 8 years ($d = 0.23$) and only then showed a relatively large decline until age 12 years (the decrease corresponded to $d = -0.73$). Between ages 12 and 18 years, only small changes were observed in the verbal domain. Also, it should be noted that the net decrease in the verbal domain (cumulative $d = -0.51$ at age 18) was much smaller compared to the mathematics domain.

Moderator Analyses

The findings reported in Table 4 suggested that there was significant heterogeneity in effect sizes for most age groups in most domains. Moreover, most I^2 values (i.e., the ratio of total

heterogeneity to total variability) were relatively large. We therefore tested whether sample characteristics moderated the effect sizes.

The variables mean year of birth and proportion of female participants were continuous and were included as such in the moderator variables. For the categorical variables, we focused on specific contrasts due to low numbers of samples in some of the categories. For country, we contrasted samples from the United States (40% in the full meta-analytic dataset) with samples from other countries (60%).[4] For ethnicity, we contrasted samples that were White/European (66%) with other samples (34%). In the moderator analyses, we also controlled for mean age of the sample, to ensure that any moderator effects of other sample characteristics were not due to a confounding with the age of the sample. Table 5 shows descriptive statistics and intercorrelations of the moderators.

Table 5
Descriptive Statistics and Intercorrelations of Moderators

Variable	*M*	*SD*	1	2	3	4	5
1. Mean age	13.45	3.83	—				
2. Mean year of birth	1986.29	9.55	−.42*	—			
3. Female (proportion)	0.54	0.33	−.03	.05	—		
4. Country[a]	0.40	0.49	.09	−.46*	.05	—	
5. Ethnicity[b]	0.66	0.47	−.06	.04	.02	−.30*	—

Note. The number of samples is $k = 143$.
[a] 1 = United States, 0 = other.
[b] 1 = White/European, 0 = other.
* $p < .05$.

Table 6 shows the results of the mixed-effects meta-regression models. We began by examining the full set of samples covering the observed age range from 5 to 28 years (see values in the left half of Table 6). Since we conducted a large number of tests (i.e., 40, as we tested five moderators in each of eight domains), we adjusted the significance level to $p < .0013$, following the Bonferroni method (i.e., dividing .05 by 40). The results showed that none of the moderators had a significant effect. Thus, the findings do not suggest that sample characteristics such as mean year of birth, gender, country (i.e., United States vs. other), and ethnicity (i.e., White/European vs. other) systematically influence mean-level change in domain-specific self-evaluations.

Given that the majority of samples, in most domains, had a mean age ranging from 10 to 16 years (Table 4), we repeated the moderator analyses for this subset of samples. These analyses may yield additional insights for two reasons. First, samples with an age 10 to 16 years are more homogeneous with regard to age, so the analyses provide information specific to this developmental period (i.e., preadolescence and adolescence). Second, by holding age relatively constant, the analyses may provide a more valid test of cohort effects on mean-level change (Baltes, Cornelius, & Nesselroade, 1979). Table 6 (values in the right half) shows the multiple regression coefficients of the moderators. Again, since we conducted a large number of tests (i.e., five moderators across seven domains; for the romantic domain, this model could not be tested), we adjusted the significance level to $p < .0014$, following the Bonferroni method (i.e., dividing .05 by 35). The results showed that none of the moderators had a significant effect in this subset of studies either.

Table 6

Mixed-Effects Meta-Regression Models for Sample Characteristics Predicting Mean-Level Change (d_{year}) in Domain-Specific Self-Evaluations

Domain and moderator	Samples with mean age from 5 to 28 years				Samples with mean age from 10 to 16 years			
	k	*B*	*SE*	*p*	*k*	*B*	*SE*	*p*
Academic	45				33			
Mean age		−.003	.006	.619		.007	.017	.694
Mean year of birth		.001	.002	.612		−.000	.003	.974
Female (proportion)		.044	.088	.619		.064	.109	.553
Country[a]		.141	.047	.003		.099	.064	.124
Ethnicity[b]		.015	.040	.712		.075	.052	.146
Appearance	61				39			
Mean age		.000	.004	.912		.005	.013	.684
Mean year of birth		.001	.002	.725		.000	.003	.963
Female (proportion)		−.082	.034	.015		−.123	.047	.010
Country[a]		−.049	.032	.132		−.050	.048	.296
Ethnicity[b]		−.004	.031	.892		−.003	.044	.940
Athletic	25				17			
Mean age		−.003	.010	.779		.028	.025	.250
Mean year of birth		.001	.004	.883		−.002	.004	.611
Female (proportion)		−.076	.130	.558		−.040	.115	.729
Country[a]		−.001	.088	.993		.057	.090	.526
Ethnicity[b]		−.040	.094	.671		.001	.090	.991
Morality	13				10			
Mean age		.009	.010	.335		.010	.030	.750
Mean year of birth		.003	.005	.554		.003	.003	.331
Female (proportion)		−.588	.356	.099		−.842	.374	.024
Country[a]		.083	.093	.373		.012	.122	.923
Ethnicity[b]		.048	.091	.597		.194	.106	.067
Romantic	7				—			
Mean age		−.044	.032	.170				
Mean year of birth		−.033	.035	.341				
Female (proportion)		.007	.087	.936				
Country[a]		−.028	.098	.773				
Ethnicity[b]		−.226	.296	.446				
Social	41				31			
Mean age		−.006	.006	.347		−.013	.017	.433
Mean year of birth		−.003	.003	.262		−.001	.003	.737
Female (proportion)		−.002	.078	.976		.024	.086	.779
Country[a]		.034	.053	.523		.045	.061	.456
Ethnicity[b]		.042	.049	.397		.041	.052	.426

Mathematics	29				20			
Mean age		−.016	.012	.182		−.002	.022	.938
Mean year of birth		−.003	.005	.472		−.007	.005	.140
Female (proportion)		−.066	.095	.488		.052	.107	.624
Country[a]		−.038	.089	.669		−.079	.096	.409
Ethnicity[b]		−.070	.074	.343		−.091	.058	.119
Verbal	14				10			
Mean age		−.009	.044	.836		.124	.056	.028
Mean year of birth		.011	.018	.554		−.003	.010	.794
Female (proportion)		−.075	.277	.787		−.056	.141	.690
Country[a]		.184	.413	.656		.146	.257	.571
Ethnicity[b]		.150	.230	.513		−.002	.096	.980

Note. Regression coefficients are unstandardized. Dash indicates that model cannot be estimated due to insufficient number of samples. d_{year} = standardized mean change *d* per year; *SE* = standard error; *k* = number of samples.
[a] 1 = United States, 0 = other.
[b] 1 = White/European, 0 = other.

Discussion

In this meta-analysis, we synthesized the available longitudinal data on mean-level change in domain-specific self-evaluations. Analyses were based on data from 143 independent samples, including 112,204 participants. The age associated with effect sizes ranged from 5 to 28 years. The results suggest that self-evaluations change systematically in many domains over the course of development from early childhood to young adulthood. The developmental trajectories of self-evaluations were positive, or at least relatively positive, in the domains of academic abilities, social acceptance, and romantic relationships. In contrast, self-evaluations showed negative, or at least relatively negative, developmental trajectories in the domains of morality, mathematics, and verbal abilities. Little mean-level change was observed in the domains of physical appearance and athletic abilities. Moderator analyses were conducted for the full set of samples and for the subset of samples between ages 10 and 16 years. The moderator analyses suggested that the pattern of mean-level change did not significantly differ by birth cohort and

gender, for samples from the United States versus other countries, and for samples with White/European versus other ethnicity.

Implications of the Findings

As reviewed in the Introduction, previous research had yielded a relatively inconsistent picture of the normative development of domain-specific self-evaluations. Nevertheless, several important themes emerged from the review of the literature. In the following, we will focus on these themes.

The transition from early to middle childhood. Previous work in this area suggested that self-evaluations decline from about age 5 to 8 years (Harter, 2006c). The general hypothesis was that self-perceptions are unrealistically positive in very young children and that, ironically, social-cognitive advances such as improvements in the use of social comparison information and social perspective taking lead to more realistic and, consequently, less positive, self-evaluations as children transition from early to middle childhood (Harter, 2006b, 2006c; Ruble et al., 1980). For most domains, this hypothesis was not supported by the present meta-analytic findings, as illustrated in Figure 2. In many domains—such as academic abilities, physical appearance, athletic abilities, and mathematics—mean-level change was small in this age period, with cumulative decreases or increases not larger than $d = 0.15$. Self-evaluations of verbal abilities even showed a slightly larger increase. However, in the domains of morality and social acceptance, self-evaluations did show a substantial decline, with effect sizes of about $d = -0.50$. Thus, it is possible that social-cognitive advances, as well as stricter expectations by parents, account for declining self-evaluations in these domains. Moreover, we note that the decline of morality self-evaluations in middle childhood could be related to negative changes that have been observed for agreeableness and honesty-humility (Klimstra, Jeronimus, Sijtsema, &

Denissen, 2020; Soto et al., 2011). Taken together, the present findings suggest that in most domains self-evaluations are relatively stable from age 5 to 8. Thus, the notion that self-evaluations typically show a substantial, or even dramatic, decline in the transition from early to middle childhood should be revised.

A low point in early adolescence? The literature suggested that mean levels of self-evaluations reach a low point in early adolescence (i.e., at age 10 to 14 years), in particular in the domains of academic abilities, social acceptance, and physical appearance (Eccles et al., 1984; Marsh, 1989; Wigfield et al., 1991). The general hypothesis was that developmental changes related to puberty and to the many environmental changes in the transition from primary to secondary school account for "hitting rock bottom" in this age period (Cantin & Boivin, 2004; Harter, 2006c; Simmons et al., 1979; Wigfield et al., 1991). In fact, self-evaluations showed a declining trend in many domains in early adolescence in this meta-analysis, including academic abilities, physical appearance, morality, mathematics, and verbal abilities, and low points emerged for the domains of physical appearance, morality, and verbal abilities. However, as illustrated in Figure 2, the declines in self-evaluations were relatively benign, with effect sizes of at most $d = -0.20$ (except for mathematics, for which the decline corresponded to $d = -0.54$). Moreover, self-evaluations in the social domains (i.e., social acceptance, romantic relationships) increased in early adolescence. Taken together, the present findings suggest that in many domains average levels of self-evaluations decline in early adolescence, but that most declines are relatively small and do not correspond to the hypothesis that the development of self-evaluations is generally characterized by reaching a low point in this age period.

However, even if there are no pronounced low points in the mean levels of self-evaluations, early adolescence might be characterized by other self-evaluation issues.

Specifically, early adolescents might show relatively large within-person fluctuations in self-evaluations over short terms such as days or weeks, for example due to all-or-none thinking and overgeneralizations (Harter, 2006c; Molloy, Ram, & Gest, 2011). In fact, research suggests that within-person fluctuations in global self-esteem decrease as early adolescents grow older and transition to later developmental stages (Meier, Orth, Denissen, & Kühnel, 2011). Also, the within-person instability of self-descriptions (Charles & Pasupathi, 2003) and affect (Larson, Moneta, Richards, & Wilson, 2002; Röcke, Li, & Smith, 2009) decreases with age. At the same time, self-concept clarity (i.e., the degree to which self-beliefs are clearly and confidently defined; Campbell et al., 1996) increases from early to middle adolescence and also throughout adulthood (Lodi-Smith & Roberts, 2010; Van Dijk et al., 2014). If self-evaluations of early adolescents fluctuate rapidly from moment to moment, or day to day, this may have contributed to the perception that the age period is especially problematic with regard to self-evaluations.

Development in late adolescence and young adulthood. The literature suggested that mean levels of self-evaluations increase from late adolescence (i.e., at about age 16 years) to young adulthood (Harter, 2006b). In these developmental stages, many personality characteristics develop in the direction of greater maturity and contribute to better functioning in social, educational, and work contexts (Bleidorn & Hopwood, 2019; Roberts & Wood, 2006; Roberts et al., 2008); consequently, positive mean-level changes can be expected for self-evaluations in these domains. Also, attachment anxiety in romantic relationships tends to decline in early adulthood, suggesting that self-evaluations in the romantic domain could become more positive (Chopik & Edelstein, 2014; Hudson, Fraley, Chopik, & Heffernan, 2015). The present findings supported this hypothesis with regard to self-evaluations of academic abilities, romantic relationships, and social acceptance. The largest effect size emerged for academic abilities (d =

0.87 from age 16 to 28 years), whereas the effect sizes were smaller for romantic relationships and social acceptance (at about $d = 0.30$). In many other domains, such as physical appearance, athletic abilities, and morality, mean-level change in self-evaluations was very small in this age period (cumulative effect sizes were not larger than $d = 0.10$). Thus, the findings suggest that positive changes in self-evaluations are restricted to general academic abilities and to the social domain. However, as noted above, these might be domains that are more relevant to the development of mature personality traits compared to other domains such as physical appearance and athletic abilities.

Similarity to the trajectory of global self-esteem. Another important theme that emerged from previous research in this area was that self-evaluations might, in some domains, show developmental trajectories that are similar to global self-esteem. As reported in the Introduction, global self-esteem tends to increase in early and middle childhood, remain at about the same level in adolescence, and increase strongly in late adolescence and young adulthood (Orth et al., 2018). Given that self-evaluations of academic abilities, physical appearance, romantic relationships, and social acceptance are substantially correlated with global self-esteem (e.g., Harris et al., 2018; Marsh, 1989; von Soest et al., 2016), there was reason to expect that the normative trajectories in these domains would be more similar to global self-esteem than the trajectories in other domains. The present meta-analytic findings supported this hypothesis for academic abilities, romantic relationships, and social acceptance. In these domains, self-evaluations generally showed a positive developmental trajectory similar to that of global self-esteem, even if there were some smaller deviations from this trend. The hypothesis was not supported for self-evaluations of physical appearance, which showed only little mean-level change across the observed age range. Supporting the hypothesis, however, was that self-

evaluations in domains for which correlations with global self-esteem are weaker—such as athletic abilities, morality, and mathematics—showed developmental trajectories that differed strongly from the trajectory of global self-esteem.

Overall, these findings are consistent with the notion that the development of global self-esteem is closely linked to that of domain-specific self-evaluations regarding academic abilities and social relationships. There are at least three theoretical models that could explain such a link (Marsh & Yeung, 1998). First, the bottom-up model proposes that domain-specific self-evaluations form the basis of global self-esteem and, consequently, causally influence the person's global self-esteem. For example, as discussed in the Introduction, sociometer theory suggests that self-perceptions of social acceptance and romantic appeal could determine the individual's feelings of self-esteem, because of the strong relevance of inclusion in close relationships and small groups in humans' evolutionary history (Leary, 2012; Leary & Baumeister, 2000; see also Kirkpatrick & Ellis, 2001; Kirkpatrick, Waugh, Valencia, & Webster, 2002). Second, the top-down model proposes that causality flows in the opposite direction; that is, from global self-esteem to domain-specific self-evaluations. Although the model does not advance domain-specific hypotheses, it is possible that the causal influence is stronger for some domains (such as social relationships) than for others (such as athletic abilities). Third, the horizontal model does not assume that global and domain-specific self-evaluations are causally linked to each other, but that other characteristics (e.g., genetic influences) account for the association between different self-evaluations. The few studies that tested the relations between global and domain-specific self-evaluations did not find evidence for the bottom-up and top-down models, thus supporting the horizontal model (Harris et al., 2018; Marsh & Yeung, 1998). However, more research on the validity of these models is needed. Robust knowledge about the

models would help to better understand the development of both global self-esteem and domain-specific self-evaluations.

In this context, it may be useful to address the interesting—and perhaps surprising—finding that the normative trajectory of academic abilities differed fundamentally from the domains of mathematics and verbal abilities. Whereas self-evaluations of general academic abilities showed a trajectory that was relatively similar to the positive trajectory of global self-esteem, self-evaluations of specific academic subjects showed a mostly negative developmental trajectory. Put differently, although mathematical and verbal abilities are considered important academic subdomains, developmental trends in self-evaluations of these subdomains cannot account for the developmental trend of general academic abilities. In our view, this discrepancy likely has to do with the fact (as reviewed in the Introduction) that general measures of academic abilities correlate much more strongly with global self-esteem than do specific measures (Esnaola et al., 2020; Marsh, 1986). A possible explanation is that children's and adolescents' self-concept of their general academic competences is only loosely related to specific experiences and grades in school subjects such as mathematics, first language, and foreign languages. Rather, it is possible that general academic self-evaluations are more closely related to the general positivity of the individual's self-concept, as indicated by measures of global self-esteem.

Moderators. A fifth theme in the literature on the development of self-evaluations is the search for moderators. Cross-sectional meta-analyses provide strong support for gender differences, such that girls and women perceive themselves more positively regarding morality and verbal abilities, whereas boys and men have more positive self-views of their athletic competences and mathematics abilities (Gentile et al., 2009; Wilgenbusch & Merrell, 1999).

However, despite gender differences in the average *levels* of self-evaluations, the few available longitudinal studies suggested that gender differences in the average *slopes* of trajectories are nonsignificant (Cole et al., 2001; Marsh, 1989; von Soest et al., 2016). The present meta-analytic findings supported the hypothesis that gender does not moderate the slopes of female and male trajectories of self-evaluations. It is important to note that this meta-analysis examines mean-level change and, consequently, is mute with regard to gender differences in the average level of trajectories. Consequently, the nonsignificant gender effect does not conflict with the findings from the cross-sectional meta-analyses cited above. The available evidence thus suggests that women and men differ in the average level of self-evaluations in some domains, but that the normative shape of developmental trajectories of self-evaluations does not differ by gender. Interestingly, a similar situation emerged in the study of global self-esteem. Despite robust evidence on the cross-sectional gender difference (Kling, Hyde, Showers, & Buswell, 1999; Zuckerman, Li, & Hall, 2016), the meta-analytic estimate of the gender difference in change was nonsignificant and virtually zero (Orth et al., 2018).

Research on generational changes in narcissism and self-esteem suggested that birth cohort membership could be a moderator of mean-level change in self-evaluations (Gentile et al., 2010; Twenge & Campbell, 2008; Twenge et al., 2008). Given sociocultural changes over the past decades, such as an increased focus on self-esteem by parents and grade inflation at school, more recent generations might show greater positivity, and more positive trajectories, in self-evaluations across many domains (Twenge & Campbell, 2010). However, the present meta-analytic findings did not support this hypothesis. In this meta-analysis, mean year of birth ranged from 1950 to 2007, so the nonsignificant moderator effects of birth cohort suggest that the shape of developmental trajectories of self-evaluations has not changed over the generations born since

the mid-20th century. For all domains of self-evaluation examined in this research, the cohort effect was nonsignificant and close to zero. Importantly, mean age was statistically controlled for in the analyses. Consequently, the effects of mean year of birth captured the unique cohort effects while holding age constant, which otherwise could have confounded the findings. Moreover, when we repeated the moderator analyses for samples with mean ages of 10 to 16 years (i.e., a subset of samples that were relatively homogeneous with regard to age), the same pattern of nonsignificant cohort effects emerged. Again, it is important to emphasize that this meta-analysis provides information only about effects on the slope but not the level of trajectories of self-evaluations. However, with regard to the average level of self-evaluations, a number of cross-sectional studies did not support the hypothesis that more recent generations differ systematically from previous generations (Trzesniewski & Donnellan, 2009, 2010; Trzesniewski et al., 2008).

The fact that the moderators tested in this meta-analysis were all nonsignificant (including the moderating effects of country and ethnicity) does not mean that there is no variability in the trajectories of self-evaluations. Rather, the heterogeneity statistics indicated that effect sizes did differ significantly across samples. Moreover, within samples, participants most certainly differed in the individual trajectories they followed. Therefore, future research should continue to test for moderators of the development of self-evaluations. Nevertheless, the present findings suggest that the pattern of findings on mean-level change is relatively robust and cannot be explained by differences in terms of gender, birth cohort, country (United States vs. other), and ethnicity (White/European vs. other).

Limitations and Strengths

A limitation of the present meta-analysis is that the samples predominantly came from Western cultural contexts; more precisely, from North America, Europe, and Australia. Only 10% of the samples were from other countries. Because of the low number of non-Western samples, the present research did not allow testing whether the pattern of findings replicates in samples from Asian, African, South American, or Central American countries. Thus, the findings of this meta-analysis essentially apply to Western countries. In future research, it would be highly desirable to more often collect data on self-evaluations in non-Western samples, to evaluate the degree to which developmental patterns generalize across cultures (Henrich et al., 2010).

Similarly, another limitation is that most of the samples included were predominantly White/European (66%). Consequently, in the moderator analyses, we focused on the contrast between White versus other samples, due to the low number of samples with predominantly Asian, Black, or Hispanic participants. Therefore, even if the moderator analyses indicated that White and other samples did not differ significantly, conclusions about the normative development of self-evaluations in specific non-White ethnic groups are not possible on the basis of the present meta-analysis. Therefore, future research should more often use ethnically diverse samples and test whether developmental patterns hold across different ethnic groups.

Unfortunately, the present meta-analysis did not allow examining age groups older than 28 years, given the lack of studies on young adulthood beyond 28 years, middle adulthood, and old age. In future research, it would be interesting to study the development of self-evaluations in adulthood. Given that global self-esteem shows substantial mean-level change across adulthood (Orth et al., 2018; for a review, see Orth & Robins, 2019), there is reason to expect that average levels of domain-specific self-evaluations also change in this period, at least in some domains.

We also note that—even if the observed age range was 5 to 28 years in the meta-analytic dataset—the number of adult samples was low for many of the domains, limiting the robustness of conclusions with regard to the normative development of self-evaluations above age 18 years.

A number of potential limitations are related to measurement of self-evaluations. For example, the primary studies included in this meta-analysis used different measures to assess self-evaluations, and some of the studies used measures that had been translated from English to other languages. Thus, the present analyses are based on the assumption that different measures, and foreign-language versions of the same measures, show convergent validity. Also, since we examined mean-level change, the analyses are based on the assumption that the measures show measurement invariance over time (Widaman, Ferrer, & Conger, 2010). Moreover, in the present meta-analysis we did not code information on the quality of measures, such as reliability, and it is possible that some of the heterogeneity in effect sizes could be explained by between-study differences in measurement quality. Nevertheless, it is important to note that many studies used established measures of the constructs, such as Harter's Self-Perception Profiles (e.g., Harter, 2012c) and Marsh's Self-Description Questionnaires (e.g., Marsh, 1990b; for a review of Harter's and Marsh's measures, see Donnellan, Trzesniewski, & Robins, 2015).

Nevertheless, an important strength of this meta-analysis is the large total number of samples that could be included in the analyses (k = 143), as well as the availability of effect sizes across a broad range of domains of self-evaluations. Moreover, we believe that the effect size measure used (i.e., the standardized mean change per year, d_{year}) significantly contributes to the validity of the meta-analytic findings. Based on age-specific estimates of d_{year}, the present analyses allowed drawing a relatively precise picture of normative development in domain-specific self-evaluations, by showing cumulative d values across the observed age range.

Furthermore, the analyses suggested that there is no publication bias in the meta-analytic dataset. We used three methods to assess publication bias, including funnel graphs (Sutton, 2009), Egger's regression test (Egger et al., 1997), and testing for differences between effect sizes from peer-reviewed journal articles versus dissertations as a category of gray literature (Ferguson & Brannick, 2012). The results of all three methods converged in suggesting that there is no evidence of publication bias in the present findings.

Conclusions

Based on longitudinal data from 143 samples with more than 110,000 individuals, this meta-analysis shows that average levels of domain-specific self-evaluations change in a systematic, normative way as individuals go through childhood, adolescence, and young adulthood. The developmental trajectories differed substantially across domains. The most positive trajectories of self-evaluations were observed with regard to general academic abilities and in the social domain. In contrast, the findings suggest that little mean-level change occurs in self-evaluations of physical appearance and athletic abilities. The most negative trajectories emerged with regard to specific academic abilities such as mathematics. The pattern of results did not differ significantly by birth cohort and gender, for samples from the United States versus other countries, and for samples with White/European versus other ethnicity, suggesting that the findings are robust and generalizable within Western cultural contexts.

Regarding the transition from early to middle childhood, the meta-analytic findings deviate from prior depictions in the literature. Thus, the notion that self-evaluations show a substantial normative decline in this age period should be revised. Similarly, the notion that self-evaluations reach a critical low point in early adolescence was not supported. Although the present findings suggest that self-evaluations typically do decline in many domains in early

adolescence, the declines were characterized by relatively small effect sizes. Nevertheless, it is important to reiterate that the present findings provide information about average developmental trajectories. Given that there are substantial individual differences in developmental trajectories of self-evaluations (Kuzucu et al., 2014; Young & Mroczek, 2003), it is likely that some children and adolescents do experience more dramatic declines and low points compared to others of their age group. Fortunately, research suggests that interventions aimed at improving domain-specific self-evaluations are effective (O'Mara, Marsh, Craven, & Debus, 2006).

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Antonishak, J. M. (2005). Predictors and sequelae of instability in adolescent peer groups. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 66*(3-B), 1752.

*Antunes, C., & Fontaine, A. M. (2007). Gender differences in the causal relation between adolescents' maths self-concept and scholastic performance. *Psychologica Belgica, 47*, 71-94. http://dx.doi.org/10.5334/pb-47-1-71/

*Arens, A. K., Marsh, H. W., Craven, R. G., Yeung, A. S., Randhawa, E., & Hasselhorn, M. (2016). Math self-concept in preschool children: Structure, achievement relations, and generalizability across gender. *Early Childhood Research Quarterly, 36*, 391-403. http://dx.doi.org/10.1016/j.ecresq.2015.12.024

Arnett, J. J. (1999). Adolescent storm and stress, reconsidered. *American Psychologist, 54*, 317-326. http://dx.doi.org/10.1037/0003-066X.54.5.317

Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist, 63*, 602-614. http://dx.doi.org/10.1037/0003-066X.63.7.602

*Asendorpf, J. B., & van Aken, M. A. G. (1994). Traits and relationship status: Stranger versus peer group inhibition and test intelligence versus peer group competence as early predictors of later self-esteem. *Child Development, 65*, 1786-1798. http://dx.doi.org/10.1111/j.1467-8624.1994.tb00849.x

Baltes, P. B., Cornelius, S. W., & Nesselroade, J. R. (1979). Cohort effects in developmental psychology. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 61-87). New York, NY: Academic Press.

*Bao, X., & Jin, K. (2015). The beneficial effect of Tai Chi on self-concept in adolescents. *International Journal of Psychology, 50*, 101-105. http://dx.doi.org/10.1002/ijop.12066

*Becker, M., & Neumann, M. (2016). Context-related changes in academic self concept development: On the long-term persistence of big-fish-little-pond effects. *Learning and Instruction, 45*, 31-39. http://dx.doi.org/10.1016/j.learninstruc.2016.06.003

*Becker, M., & Neumann, M. (2018). Longitudinal big-fish-little-pond effects on academic self-concept development during the transition from elementary to secondary schooling. *Journal of Educational Psychology, 110*, 882-897. http://dx.doi.org/10.1037/edu0000233

*Becker, M., Neumann, M., Tetzner, J., Böse, S., Knoppick, H., Maaz, K., . . . Lehmann, R. (2014). Is early ability grouping good for high-achieving students' psychosocial development? Effects of the transition into academically selective schools. *Journal of Educational Psychology, 106*, 555-568. http://dx.doi.org/10.1037/a0035425

*Bellmore, A. D., & Cillessen, A. H. N. (2006). Reciprocal influences of victimization, perceived social preference, and self-concept in adolescence. *Self and Identity, 5*, 209-229. http://dx.doi.org/10.1080/15298860600636647

Bleidorn, W., & Hopwood, C. J. (2019). Stability and change in personality traits over the lifespan. In D. P. McAdams, R. L. Shiner, & J. L. Tackett (Eds.), *Handbook of personality development* (pp. 237-252). New York, NY: Guilford.

*Bodkin-Andrews, G. H., O'Rourke, V., Dillon, A., Craven, R. G., & Yeung, A. S. (2012). Engaging the disengaged? A longitudinal analysis of the relations between indigenous

and non-indigenous Australian students' academic self-concept and disengagement. *Journal of Cognitive Education and Psychology, 11*, 179-195. http://dx.doi.org/10.1891/1945-8959.11.2.179

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.

*Bornholt, L. J., & Piccolo, A. (2005). Individuality, belonging, and children's self concepts: A motivational spiral model of self-evaluation, performance, and participation in physical activities. *Applied Psychology: An International Review, 54*, 515-536. http://dx.doi.org/10.1111/j.1464-0597.2005.00223.x

*Bosacki, S. L. (2015). Children's theory of mind, self-perceptions, and peer relations: A longitudinal study. *Infant and Child Development, 24*, 175-188. http://dx.doi.org/10.1002/icd.1878

*Brown, K. M., McMahon, R. P., Biro, F. M., Crawford, P., Schreiber, G. B., Similo, S. L., . . . Striegel-Moore, R. (1998). Changes in self-esteem in Black and White girls between the ages of 9 and 14 years. *Journal of Adolescent Health, 23*, 7-19. http://dx.doi.org/10.1016/S1054-139X(97)00238-3

*Cairns, E., McWhirter, L., Duffy, U., & Barry, R. (1990). The stability of self-concept in late adolescence: Gender and situational effects. *Personality and Individual Differences, 11*, 937-944. http://dx.doi.org/10.1016/0191-8869(90)90275-V

Campbell, J. D., Trapnell, P. D., Heine, S. J., Katz, I. M., Lavallee, L. F., & Lehman, D. R. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of Personality and Social Psychology, 70*, 141-156. http://dx.doi.org/10.1037/0022-3514.70.1.141

*Cantin, S., & Boivin, M. (2004). Change and stability in children's social network and self-perceptions during transition from elementary to junior high school. *International Journal of Behavioral Development, 28*, 561-570. http://dx.doi.org/10.1080/01650250444000289

*Cast, A. D., & Cantwell, A. M. (2007). Identity change in newly married couples: Effects of positive and negative feedback. *Social Psychology Quarterly, 70*, 172-185. http://dx.doi.org/10.1177/019027250707000206

*Chapman, J. W., & Tunmer, W. E. (1997). A longitudinal study of beginning reading achievement and reading self-concept. *British Journal of Educational Psychology, 67*, 279-291. http://dx.doi.org/10.1111/j.2044-8279.1997.tb01244.x

Charles, S. T., & Pasupathi, M. (2003). Age-related patterns of variability in self-descriptions: Implications for everyday affective experience. *Psychology and Aging, 18*, 524-536. http://dx.doi.org/10.1037/0882-7974.18.3.524

*Chen, H., & Jackson, T. (2009). Predictors of changes in weight esteem among Mainland Chinese adolescents: A longitudinal analysis. *Developmental Psychology, 45*, 1618-1629. http://dx.doi.org/10.1037/a0016820

*Chen, X., He, Y., & Li, D. (2004). Self-perceptions of social competence and self-worth in Chinese children: Relations with social and school performance. *Social Development, 13*, 570-589. http://dx.doi.org/10.1111/j.1467-9507.2004.00284.x

Chopik, W. J., & Edelstein, R. S. (2014). Age differences in romantic attachment around the world. *Social Psychological and Personality Science, 5*, 892-900. http://dx.doi.org/10.1177/1948550614538460

Chung, J. M., Robins, R. W., Trzesniewski, K. H., Noftle, E. E., Roberts, B. W., & Widaman, K. F. (2014). Continuity and change in self-esteem during emerging adulthood. *Journal of Personality and Social Psychology, 106*, 469-483. http://dx.doi.org/10.1037/a0035135

*Clark, L., & Tiggemann, M. (2008). Sociocultural and individual psychological predictors of body image in young girls: A prospective study. *Developmental Psychology, 44*, 1124-1134. http://dx.doi.org/10.1037/0012-1649.44.4.1124

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cole, D. A., Maxwell, S. E., Martin, J. M., Peeke, L. G., Seroczynski, A. D., Tram, J. M., . . . Maschman, T. (2001). The development of multiple domains of child and adolescent self-concept: A cohort sequential longitudinal design. *Child Development, 72*, 1723-1746. http://dx.doi.org/10.1111/1467-8624.00375

*Crocker, P. R. E., Sabiston, C. M., Kowalski, K. C., McDonough, M. H., & Kowalski, N. (2006). Longitudinal assessment of the relationship between physical self-concept and health-related behavior and emotion in adolescent girls. *Journal of Applied Sport Psychology, 18*, 185-200. http://dx.doi.org/10.1080/10413200600830257

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*, 7-29. http://dx.doi.org/10.1177/0956797613504966

*Dapp, L. C., & Roebers, C. M. (2018). Self-concept in kindergarten and first grade children: A longitudinal study on structure, development, and relation to achievement. *Psychology, 9*, 1605-1629. http://dx.doi.org/10.4236/psych.2018.97097

*Davis-Kean, P. E., Huesmann, L. R., Jager, J., Collins, W. A., Bates, J. E., & Lansford, J. E. (2008). Changes in the relation of self-efficacy beliefs and behaviors across development. *Child Development, 79*, 1257-1269. http://dx.doi.org/10.1111/j.1467-8624.2008.01187.x

*Davison, K. K., Schmalz, D. L., Young, L. M., & Birch, L. L. (2008). Overweight girls who internalize fat stereotypes report low psychosocial well-being. *Obesity, 16*, S30-S38. http://dx.doi.org/10.1038/oby.2008.451

*Davison, K. K., Werder, J. L., Trost, S. G., Baker, B. L., & Birch, L. L. (2007). Why are early maturing girls less active? Links between pubertal development, psychological well-being, and physical activity among girls at ages 11 and 13. *Social Science and Medicine, 64*, 2391-2404. http://dx.doi.org/10.1016/j.socscimed.2007.02.033

*De Neve, J. E., & Oswald, A. J. (2012). Estimating the influence of life satisfaction and positive affect on later income using sibling fixed effects. *Proceedings of the National Academy of Sciences, 109*, 19953-19958. http://dx.doi.org/10.1073/pnas.1211437109

Denissen, J. J. A., Penke, L., Schmitt, D. P., & Van Aken, M. A. G. (2008). Self-esteem reactions to social interactions: Evidence for sociometer mechanisms across days, people, and nations. *Journal of Personality and Social Psychology, 95*, 181-196. http://dx.doi.org/10.1037/0022-3514.95.1.181

*Denny, M. S. (2011). The relationships among internalizing symptomatology in kindergarten and later self-concept and competence. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 72*(4-A), 1170.

*Dickhäuser, O., Janke, S., Praetorius, A. K., & Dresel, M. (2017). The effects of teachers' reference norm orientations on students' implicit theories and academic self-concepts. *Zeitschrift für Pädagogische Psychologie, 31*, 205-219. http://dx.doi.org/10.1024/1010-0652/a000208

*Diemer, M. A., Marchand, A. D., McKellar, S. E., & Malanchuk, O. (2016). Promotive and corrosive factors in African American students' math beliefs and achievement. *Journal of Youth and Adolescence, 45*, 1208-1225. http://dx.doi.org/10.1007/s10964-016-0439-9

*Dohnt, H., & Tiggemann, M. (2006). The contribution of peer and media influences to the development of body satisfaction and self-esteem in young girls: A prospective study. *Developmental Psychology, 42*, 929-936. http://dx.doi.org/10.1037/0012-1649.42.5.929

*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*, 453-472. http://dx.doi.org/10.1016/j.jrp.2006.06.004

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, UK: Elsevier.

*Dufner, M., Reitz, A. K., & Zander, L. (2015). Antecedents, consequences, and mechanisms: On the longitudinal interplay between academic self-enhancement and psychological adjustment. *Journal of Personality, 83*, 511-522. http://dx.doi.org/10.1111/jopy.12128

Eccles, J. S., Midgley, C., & Adler, T. F. (1984). Grade-related changes in the school environment: Effects on achievement motivation. In J. G. Nicholls (Ed.), *The development of achievement motivation* (pp. 283-331). Greenwich, CT: JAI Press.

Eccles, J. S., Wigfield, A., Flanagan, C. A., Miller, C., Reuman, D. A., & Yee, D. (1989). Self-concepts, domain values, and self-esteem: Relations and changes at early adolescence. *Journal of Personality, 57*, 283-310. http://dx.doi.org/10.1111/j.1467-6494.1989.tb00484.x

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629-634. http://dx.doi.org/10.1136/bmj.315.7109.629

*Eisenberg, M. E., Neumark-Sztainer, D., Haines, J., & Wall, M. (2006). Weight-teasing and emotional well-being in adolescents: Longitudinal findings from Project EAT. *Journal of Adolescent Health, 38*, 675-683. http://dx.doi.org/10.1016/j.jadohealth.2005.07.002

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*, 607-619. http://dx.doi.org/10.1037/a0024299

Esnaola, I., Sesé, A., Antonio-Agirre, I., & Azpiazu, L. (2020). The development of multiple self-concept dimensions during adolescence. *Journal of Research on Adolescence, 30*, 100-114. http://dx.doi.org/10.1111/jora.12451

*Fenzel, L. M. (2000). Prospective study of changes in global self-worth and strain during the transition to middle school. *Journal of Early Adolescence, 20*, 93-116. http://dx.doi.org/10.1177/0272431600020001005

Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods, 17*, 120-128. http://dx.doi.org/10.1037/a0024445

*Ferreiro, F., Seoane, G., & Senra, C. (2012). Gender-related risk and protective factors for depressive symptoms and disordered eating in adolescence: A 4-year longitudinal study. *Journal of Youth and Adolescence, 41*, 607-622. http://dx.doi.org/10.1007/s10964-011-9718-7

Fraley, R. C., & Marks, M. J. (2007). The null hypothesis significance-testing debate and its implications for personality research. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 149-169). New York, NY: Guilford.

*Frisén, A., Lunde, C., & Berg, A. I. (2015). Developmental patterns in body esteem from late childhood to young adulthood: A growth curve analysis. *European Journal of Developmental Psychology, 12*, 99-115. http://dx.doi.org/10.1080/17405629.2014.951033

Gentile, B., Grabe, S., Dolan-Pascoe, B., Twenge, J. M., Wells, B. E., & Maitino, A. (2009). Gender differences in domain-specific self-esteem: A meta-analysis. *Review of General Psychology, 13*, 34-45. http://dx.doi.org/10.1037/a0013689

Gentile, B., Twenge, J. M., & Campbell, W. K. (2010). Birth cohort differences in self-esteem, 1988-2008: A cross-temporal meta-analysis. *Review of General Psychology, 14*, 261-268. http://dx.doi.org/10.1037/a0019919

*Gest, S. D., Domitrovich, C. E., & Welsh, J. A. (2005). Peer academic reputation in elementary school: Associations with changes in self-concept and academic skills. *Journal of Educational Psychology, 97*, 337-346. http://dx.doi.org/10.1037/0022-0663.97.3.337

*Gestsdottir, S., Svansdottir, E., Ommundsen, Y., Arnarsson, A., Arngrimsson, S., Sveinsson, T., & Johannsson, E. (2016). Do aerobic fitness and self-reported fitness in adolescence differently predict body image in young adulthood? An eight year follow-up study. *Mental Health and Physical Activity, 10*, 40-47. http://dx.doi.org/10.1016/j.mhpa.2015.12.001

*Gniewosz, B., Eccles, J. S., & Noack, P. (2012). Secondary school transition and the use of different sources of information for the construction of the academic self-concept. *Social Development, 21*, 537-557. http://dx.doi.org/10.1111/j.1467-9507.2011.00635.x

Göllner, R., Roberts, B. W., Damian, R. I., Lüdtke, O., Jonkmann, K., & Trautwein, U. (2017). Whose "storm and stress" is it? Parent and child reports of personality development in the transition to early adolescence. *Journal of Personality, 85*, 376-387. http://dx.doi.org/10.1111/jopy.12246

Gore, J. S., & Cross, S. E. (2014). Who am I becoming? A theoretical framework for understanding self-concept change. *Self and Identity, 13*, 740-764. http://dx.doi.org/10.1080/15298868.2014.933712

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin, 82*, 1-20. http://dx.doi.org/10.1037/h0076157

*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*, 1563-1577. http://dx.doi.org/10.1037/dev0000147

*Guay, F., Boivin, M., & Hodges, E. V. E. (1999). Predicting change in academic achievement: A model of peer experiences and self-system processes. *Journal of Educational Psychology, 91*, 105-115. http://dx.doi.org/10.1037/0022-0663.91.1.105

Guérin, E., Goldfield, G., & Prud'homme, D. (2017). Trajectories of mood and stress and relationships with protective factors during the transition to menopause: Results using latent class growth modeling in a Canadian cohort. *Archives of Women's Mental Health, 20*, 733-745. http://dx.doi.org/10.1007/s00737-017-0755-4

*Guo, J., Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2015). Directionality of the associations of high school expectancy-value, aspirations, and attainment: A longitudinal study. *American Educational Research Journal, 52*, 371-402. http://dx.doi.org/10.3102/0002831214565786

*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*, 4-16. http://dx.doi.org/10.1177/0165025416679744

Harter, S. (1982). The Perceived Competence Scale for Children. *Child Development, 53*, 87-97. http://dx.doi.org/10.2307/1129640

Harter, S. (2006a). The development of self-esteem. In M. H. Kernis (Ed.), *Self-esteem issues and answers: A sourcebook of current perspectives* (pp. 144-150). New York, NY: Psychology Press.

Harter, S. (2006b). Developmental and individual difference perspectives on self-esteem. In D. K. Mroczek & T. D. Little (Eds.), *Handbook of personality development* (pp. 311-334). Mahwah, NJ: Erlbaum.

Harter, S. (2006c). The self. In N. Eisenberg (Ed.), *Handbook of child psychology (6th ed.). Volume 3: Social, emotional, and personality development* (pp. 505-570). Hoboken, NJ: Wiley.

Harter, S. (2012a). *The construction of the self: Developmental and sociocultural foundations*. New York, NY: Guilford.

Harter, S. (2012b). *Self-Perception Profile for Adolescents: Manual and questionnaires*. Denver, CO: University of Denver.

Harter, S. (2012c). *Self-Perception Profile for Children: Manual and questionnaires (Grades 3-8)*. Denver, CO: University of Denver.

Harter, S., & Pike, R. (1984). The Pictorial Scale of Perceived Competence and Social Acceptance for Young Children. *Child Development, 55*, 1969-1982. http://dx.doi.org/10.2307/1129772

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

Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review, 106*, 766-794. http://dx.doi.org/10.1037/0033-295X.106.4.766

*Helmke, A., & van Aken, M. A. G. (1995). The causal ordering of academic achievement and self-concept of ability during elementary school: A longitudinal study. *Journal of Educational Psychology, 87*, 624-637. http://dx.doi.org/10.1037/0022-0663.87.4.624

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61-83. http://dx.doi.org/10.1017/S0140525X0999152X

*Hoge, D. R., Smit, E. K., & Hanson, S. L. (1990). School experiences predicting changes in self-esteem of sixth- and seventh-grade students. *Journal of Educational Psychology, 82*, 117-127. http://dx.doi.org/10.1037/0022-0663.82.1.117

*Hoglund, C. L. (1995). Longitudinal study of gender differences in global and domain-specific self-esteem in children. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 56*(5-B), 2898.

Hollenstein, T., & Lougheed, J. P. (2013). Beyond storm and stress: Typicality, transactions, timing, and temperament to account for adolescent change. *American Psychologist, 68*, 444-454. http://dx.doi.org/10.1037/a0033586

Huang, C. (2010). Mean-level change in self-esteem from childhood through adulthood: Meta-analysis of longitudinal studies. *Review of General Psychology, 14*, 251-260. http://dx.doi.org/10.1037/a0020543

Hudson, N. W., Fraley, R. C., Chopik, W. J., & Heffernan, M. E. (2015). Not all attachment relationships develop alike: Normative cross-sectional age trajectories in attachment to romantic partners, best friends, and parents. *Journal of Research in Personality, 59*, 44-55. http://dx.doi.org/10.1016/j.jrp.2015.10.001

*Impett, E. A., Henson, J. M., Breines, J. G., Schooler, D., & Tolman, D. L. (2011). Embodiment feels better: Girls' body objectification and well-being across adolescence. *Psychology of Women Quarterly, 35*, 46-58. http://dx.doi.org/10.1177/0361684310391641

*Ireson, J., & Hallam, S. (2009). Academic self-concepts in adolescence: Relations with achievement and ability grouping in schools. *Learning and Instruction, 19*, 201-213. http://dx.doi.org/10.1016/j.learninstruc.2008.04.001

*Keel, P. K., Fulkerson, J. A., & Leon, G. R. (1997). Disordered eating precursors in pre- and early adolescent girls and boys. *Journal of Youth and Adolescence, 26*, 203-216. http://dx.doi.org/10.1023/A:1024504615742

*Kimber, B., Sandell, R., & Bremberg, S. (2008). Social and emotional training in Swedish classrooms for the promotion of mental health: Results from an effectiveness study in

Sweden. *Health Promotion International, 23*, 134-143. http://dx.doi.org/10.1093/heapro/dam046

*King, R. B., & McInerney, D. M. (2014). Mapping changes in students' English and math self-concepts: A latent growth model study. *Educational Psychology, 34*, 581-597. http://dx.doi.org/10.1080/01443410.2014.909009

Kirkpatrick, L. A., & Ellis, B. J. (2001). An evolutionary-psychological approach to self-esteem: Multiple domains and multiple functions. In G. J. O. Fletcher & M. S. Clark (Eds.), *Blackwell handbook of social psychology: Interpersonal processes* (pp. 411-436). Malden, MA: Blackwell.

Kirkpatrick, L. A., Waugh, C. E., Valencia, A., & Webster, G. D. (2002). The functional domain specificity of self-esteem and the differential prediction of aggression. *Journal of Personality and Social Psychology, 82*, 756-767. http://dx.doi.org/10.1037/0022-3514.82.5.756

Klimstra, T. A., Jeronimus, B. F., Sijtsema, J. J., & Denissen, J. J. A. (2020). The unfolding dark side: Age trends in dark personality features. *Journal of Research in Personality, 85*, 103915. http://dx.doi.org/10.1016/j.jrp.2020.103915

Kling, K. C., Hyde, J. S., Showers, C. J., & Buswell, B. N. (1999). Gender differences in self-esteem: A meta-analysis. *Psychological Bulletin, 125*, 470-500. http://dx.doi.org/10.1037/0033-2909.125.4.470

*Kuzucu, Y., Bontempo, D. E., Hofer, S. M., Stallings, M. C., & Piccinin, A. M. (2014). Developmental change and time-specific variation in global and specific aspects of self-concept in adolescence and association with depressive symptoms. *Journal of Early Adolescence, 34*, 638-666. http://dx.doi.org/10.1177/0272431613507498

*LaGrange, B., Cole, D. A., Dallaire, D. H., Ciesla, J. A., Pineda, A. Q., Truss, A. E., & Folmer, A. (2008). Developmental changes in depressive cognitions: A longitudinal evaluation of the Cognitive Triad Inventory for Children. *Psychological Assessment, 20*, 217-226. http://dx.doi.org/10.1037/1040-3590.20.3.217

*Lamote, C., Pinxten, M., Van den Noortgate, W., & Van Damme, J. (2014). Is the cure worse than the disease? A longitudinal study on the effect of grade retention in secondary education on achievement and academic self-concept. *Educational Studies, 40*, 496-514. http://dx.doi.org/10.1080/03055698.2014.936828

Larson, R. W., Moneta, G., Richards, M. H., & Wilson, S. (2002). Continuity, stability, and change in daily emotional experience across adolescence. *Child Development, 73*, 1151-1165. http://dx.doi.org/10.1111/1467-8624.00464

*Le Bars, H., Gernigon, C., & Ninot, G. (2009). Personal and contextual determinants of elite young athletes' persistence or dropping out over time. *Scandinavian Journal of Medicine and Science in Sports, 19*, 274-285. http://dx.doi.org/10.1111/j.1600-0838.2008.00786.x

Leary, M. R. (2012). Sociometer theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 141-159). Thousand Oaks, CA: Sage.

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* (Vol. 32, pp. 1-62). San Diego, CA: Academic Press.

Leary, M. R., Haupt, A. L., Strausser, K. S., & Chokel, J. T. (1998). Calibrating the sociometer: The relationship between interpersonal appraisals and state self-esteem. *Journal of*

Personality and Social Psychology, 74, 1290-1299. http://dx.doi.org/10.1037/0022-3514.74.5.1290

*Liu, W. C., Wang, C. K. J., & Parkins, E. J. (2005). A longitudinal study of students' academic self-concept in a streamed setting: The Singapore context. *British Journal of Educational Psychology, 75*, 567-586. http://dx.doi.org/10.1348/000709905X42239

Lodi-Smith, J., & Roberts, B. W. (2010). Getting to know me: Social role experiences and age differences in self-concept clarity during adulthood. *Journal of Personality, 78*, 1383-1410. http://dx.doi.org/10.1111/j.1467-6494.2010.00655.x

Lucas, R. E., & Donnellan, M. B. (2011). Personality development across the life span: Longitudinal analyses with a national sample from Germany. *Journal of Personality and Social Psychology, 101*, 847-861. http://dx.doi.org/10.1037/a0024298

*Lüdtke, O., Köller, O., Marsh, H. W., & Trautwein, U. (2005). Teacher frame of reference and the big-fish–little-pond effect. *Contemporary Educational Psychology, 30*, 263-285. http://dx.doi.org/10.1016/j.cedpsych.2004.10.002

*Luszczynska, A., & Abraham, C. (2012). Reciprocal relationships between three aspects of physical self-concept, vigorous physical activity, and lung function: A longitudinal study among late adolescents. *Psychology of Sport and Exercise, 13*, 640-648. http://dx.doi.org/10.1016/j.psychsport.2012.04.003

*Mantzicopoulos, P. (2006). Younger children's changing self-concepts: Boys and girls from preschool through second grade. *Journal of Genetic Psychology, 167*, 289-308. http://dx.doi.org/10.3200/GNTP.167.3.289-308

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224-253. http://dx.doi.org/10.1037/0033-295X.98.2.224

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

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

Marsh, H. W. (1990a). A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification. *Educational Psychology Review, 2*, 77-172. http://dx.doi.org/10.1007/BF01322177

Marsh, H. W. (1990b). *Self Description Questionnaire (SDQ) II: Manual*. Campbelltown, New South Wales, Australia: University of Western Sydney, SELF Research Centre.

Marsh, H. W., Barnes, J., Cairns, L., & Tidman, M. (1984). Self-Description Questionnaire: Age and sex effects in the structure and level of self-concept for preadolescent children. *Journal of Educational Psychology, 76*, 940-956. http://dx.doi.org/10.1037/0022-0663.76.5.940

*Marsh, H. W., Craven, R., & Debus, R. (1998). Structure, stability, and development of young children's self-concepts: A multicohort-multioccasion study. *Child Development, 69*, 1030-1053. http://dx.doi.org/10.2307/1132361

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*, 133-163. http://dx.doi.org/10.1111/j.1745-6916.2006.00010.x

*Marsh, H. W., Ellis, L. A., Parada, R. H., Richards, G., & Heubeck, B. G. (2005). A short version of the Self Description Questionnaire II: Operationalizing criteria for short-form evaluation with new applications of confirmatory factor analysis. *Psychological Assessment, 17*, 81-102. http://dx.doi.org/10.1037/1040-3590.17.1.181

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*, 153-174. http://dx.doi.org/10.1111/j.1745-3984.1984.tb00227.x

Marsh, H. W., & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. *Journal of Educational Psychology, 89*, 41-54. http://dx.doi.org/10.1037/0022-0663.89.1.41

Marsh, H. W., & Yeung, A. S. (1998). Top-down, bottom-up, and horizontal models: The direction of causality in multidimensional, hierarchical self-concept models. *Journal of Personality and Social Psychology, 75*, 509-527. http://dx.doi.org/10.1037/0022-3514.75.2.509

*McGrath, E. P., & Repetti, R. L. (2002). A longitudinal study of children's depressive symptoms, self-perceptions, and cognitive distortions about the self. *Journal of Abnormal Psychology, 111*, 77-87. http://dx.doi.org/10.1037//021-843x.111.1.77

*McGuire, S., Manke, B., Saudino, K. J., Reiss, D., Hetherington, E. M., & Plomin, R. (1999). Perceived competence and self-worth during adolescence: A longitudinal behavioral genetic study. *Child Development, 70*, 1283-1296. http://dx.doi.org/10.1111/1467-8624.00094

*McKinley, N. M. (2006). Longitudinal gender differences in objectified body consciousness and weight-related attitudes and behaviors: Cultural and developmental contexts in the transition from college. *Sex Roles, 54*, 159-173. http://dx.doi.org/10.1007/s11199-006-9335-1

McLeod, B. D., & Weisz, J. R. (2004). Using dissertations to examine potential bias in child and adolescent clinical trials. *Journal of Consulting and Clinical Psychology, 72*, 235-251. http://dx.doi.org/10.1037/0022-006x.72.2.235

*McLeod, J. D., & Owens, T. J. (2004). Psychological well-being in the early life course: Variations by socioeconomic status, gender, and race/ethnicity. *Social Psychology Quarterly, 67*, 257-278. http://dx.doi.org/10.1177/019027250406700303

Meier, L. L., Orth, U., Denissen, J. J. A., & Kühnel, A. (2011). Age differences in instability, contingency, and level of self-esteem across the life span. *Journal of Research in Personality, 45*, 604-612. http://dx.doi.org/10.1016/j.jrp.2011.08.008

Mendelson, B. K., Mendelson, M. J., & White, D. R. (2001). Body-Esteem Scale for Adolescents and Adults. *Journal of Personality Assessment, 76*, 90-106. http://dx.doi.org/10.1207/S15327752JPA7601_6

Messer, B., & Harter, S. (2012). The Self-Perception Profile for Adults: Manual and questionnaires. In. Colorado, USA: University of Denver.

*Modecki, K. L., Neira, C. B., & Barber, B. L. (2018). Finding what fits: Breadth of participation at the transition to high school mitigates declines in self-concept. *Developmental Psychology, 54*, 1954-1970. http://dx.doi.org/10.1037/dev0000570

*Möller, J., Retelsdorf, J., Köller, O., & Marsh, H. W. (2011). The reciprocal internal/external frame of reference model: An integration of models of relations between academic

achievement and self-concept. *American Educational Research Journal, 48*, 1315-1346. http://dx.doi.org/10.3102/0002831211419649

*Möller, J., Zimmermann, F., & Köller, O. (2014). The reciprocal internal/external frame of reference model using grades and test scores. *British Journal of Educational Psychology, 84*, 591-611. http://dx.doi.org/10.1111/bjep.12047

Molloy, L. E., Ram, N., & Gest, S. D. (2011). The storm and stress (or calm) of early adolescent self-concepts: Within- and between-subjects variability. *Developmental Psychology, 47*, 1589-1607. http://dx.doi.org/10.1037/a0025413

*Morgan, P. J., Saunders, K. L., & Lubans, D. R. (2012). Improving physical self-perception in adolescent boys from disadvantaged schools: Psychological outcomes from the Physical Activity Leaders randomized controlled trial. *Pediatric Obesity, 7*, e27-e32. http://dx.doi.org/10.1111/j.2047-6310.2012.00050.x

*Morin, A. J. S., Maiano, 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*, 157-201. http://dx.doi.org/10.1080/00273171.2010.546731

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*, 105-125. http://dx.doi.org/10.1037/1082-989X.7.1.105

*Mössle, T., & Rehbein, F. (2013). Predictors of problematic video game usage in childhood and adolescence. *Sucht, 59*, 153-164. http://dx.doi.org/10.1024/0939-5911.a000247

*Mullins, E. R. (1997). Changes in young adolescents' self-perceptions across the transition from elementary to middle school. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 58*(12-A), 4561.

Mund, M., & Neyer, F. J. (2014). Treating personality-relationship transactions with respect: Narrow facets, advanced models, and extended time frames. *Journal of Personality and Social Psychology, 107*, 352-368. http://dx.doi.org/10.1037/a0036719

Murray, S. L., Griffin, D. W., Rose, P., & Bellavia, G. M. (2003). Calibrating the sociometer: The relational contingencies of self-esteem. *Journal of Personality and Social Psychology, 85*, 63-84. http://dx.doi.org/10.1037/0022-3514.85.1.63

*Musu-Gillette, L. E., Wigfield, A., Harring, J. R., & Eccles, J. S. (2015). Trajectories of change in students' self-concepts of ability and values in math and college major choice. *Educational Research and Evaluation, 21*, 343-370. http://dx.doi.org/10.1080/13803611.2015.1057161

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

Neemann, J., & Harter, S. (2012). *The Self-Perception Profile for College Students: Manual and questionnaires*. Colorado, USA: University of Denver.

*Nelson, S. C., Kling, J., Wängqvist, M., Frisén, A., & Syed, M. (2018). Identity and the body: Trajectories of body esteem from adolescence to emerging adulthood. *Developmental Psychology, 54*, 1159-1171. http://dx.doi.org/10.1037/dev0000435

*Newman, R. S. (1984). Children's achievement and self-evaluations in mathematics: A longitudinal study. *Journal of Educational Psychology, 76*, 857-873. http://dx.doi.org/10.1037/0022-0663.76.5.857

Neyer, F. J., & Asendorpf, J. B. (2001). Personality-relationship transaction in young adulthood. *Journal of Personality and Social Psychology, 81*, 1190-1204. http://dx.doi.org/10.1037/0022-3514.81.6.1190

*Niepel, C., Brunner, M., & Preckel, F. (2014). The longitudinal interplay of students' academic self-concepts and achievements within and across domains: Replicating and extending the reciprocal internal/external frame of reference model. *Journal of Educational Psychology, 106*, 1170-1191. http://dx.doi.org/10.1037/a0036307

Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big Five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology, 93*, 116-130. http://dx.doi.org/10.1037/0022-3514.93.1.116

*Nogueira Avelar e Silva, R., van de Bongardt, D., Baams, L., & Raat, H. (2018). Bidirectional associations between adolescents' sexual behaviors and psychological well-being. *Journal of Adolescent Health, 62*, 63-71. http://dx.doi.org/10.1016/j.jadohealth.2017.08.008

*O'Dea, J. A. (2006). Self-concept, self-esteem and body weight in adolescent females: A three-year longitudinal study. *Journal of Health Psychology, 11*, 599-611. http://dx.doi.org/10.1177/1359105306065020

*O'Dea, J. A., & Abraham, S. (2000). Improving the body image, eating attitudes, and behaviors of young male and female adolescents: A new educational approach that focuses on self-

esteem. *International Journal of Eating Disorders, 28*, 43-57. http://dx.doi.org/10.1002/(SICI)1098-108X(200007)28:1<43::AID-EAT6>3.0.CO;2-D

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*, 181-206. http://dx.doi.org/10.1207/s15326985ep4103_4

*Ohannessian, C. M., Lerner, R. M., von Eye, A., & Lerner, J. V. (1996). Direct and indirect relations between perceived parental acceptance, perceptions of the self, and emotional adjustment during early adolescence. *Family and Consumer Sciences Research Journal, 25*, 159-183. http://dx.doi.org/10.1177/1077727X960252004

*Ohannessian, C. M., Vannucci, A., Lincoln, C. R., Flannery, K. M., & Trinh, A. (2019). Self-competence and depressive symptoms in middle-late adolescence: Disentangling the direction of effect. *Journal of Research on Adolescence, 29*, 736-751. http://dx.doi.org/10.1111/jora.12412

Orth, U. (2017). The lifespan development of self-esteem. In J. Specht (Ed.), *Personality development across the lifespan* (pp. 181-195). London, UK: Elsevier.

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*, 1045-1080. http://dx.doi.org/10.1037/bul0000161

Orth, U., & Robins, R. W. (2014). The development of self-esteem. *Current Directions in Psychological Science, 23*, 381-387. http://dx.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: Guilford.

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*, 1271-1288. http://dx.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*, 622-633. http://dx.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*, 645-658. http://dx.doi.org/10.1037/a0018769

*Pomerantz, E. M., & Dong, W. (2006). Effects of mothers' perceptions of children's competence: The moderating role of mothers' theories of competence. *Developmental Psychology, 42*, 950-961. http://dx.doi.org/10.1037/0012-1649.42.5.950

*Preckel, F., Niepel, C., Schneider, M., & Brunner, M. (2013). Self-concept in adolescence: A longitudinal study on reciprocal effects of self-perceptions in academic and social domains. *Journal of Adolescence, 36*, 1165-1175. http://dx.doi.org/10.1016/j.adolescence.2013.09.001

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org.

*Radin, S. M. (2005). The development of drinking in urban American Indian adolescents: A longitudinal examination of self-derogation theory. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *66*(12-B), 6933.

Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 295-315). New York, NY: Russell Sage Foundation.

*Raustorp, A., Archer, T., Svensson, K., Perlinger, T., & Alricsson, M. (2009). Physical self-esteem, a five year follow-up study on Swedish adolescents. *International Journal of Adolescent Medicine and Health*, *21*, 497-507. http://dx.doi.org/10.1515/IJAMH.2009.21.4.497

Roberts, B. W., Caspi, A., & Moffitt, T. E. (2003). Work experiences and personality development in young adulthood. *Journal of Personality and Social Psychology*, *84*, 582-593. http://dx.doi.org/10.1037/0022-3514.84.3.582

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*, 1-25. http://dx.doi.org/10.1037/0033-2909.132.1.1

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 personality development* (pp. 11-39). Mahwah, NJ: Erlbaum.

Roberts, B. W., Wood, D., & Caspi, A. (2008). The development of personality traits in adulthood. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 375-398). New York, NY: Guilford.

Robins, R. W., & Trzesniewski, K. H. (2005). Self-esteem development across the lifespan. *Current Directions in Psychological Science, 14*, 158-162. http://dx.doi.org/10.1111/j.0963-7214.2005.00353.x

Robins, R. W., Trzesniewski, K. H., Tracy, J. L., Gosling, S. D., & Potter, J. (2002). Global self-esteem across the life span. *Psychology and Aging, 17*, 423-434. http://dx.doi.org/10.1037/0882-7974.17.3.423

Röcke, C., Li, S. C., & Smith, J. (2009). Intraindividual variability in positive and negative affect over 45 days: Do older adults fluctuate less than young adults? *Psychology and Aging, 24*, 863-878. http://dx.doi.org/10.1037/a0016276

*Roebers, C. M., Cimeli, P., Röthlisberger, M., & Neuenschwander, R. (2012). Executive functioning, metacognition, and self-perceived competence in elementary school children: An explorative study on their interrelations and their role for school achievement. *Metacognition and Learning, 7*, 151-173. http://dx.doi.org/10.1007/s11409-012-9089-9

Ruble, D. N., Boggiano, A. K., Feldman, N. S., & Loebl, J. H. (1980). Developmental analysis of the role of social comparison in self-evaluation. *Developmental Psychology, 16*, 105-115. http://dx.doi.org/10.1037/0012-1649.16.2.105

*Sallquist, J., Eisenberg, N., French, D. C., Purwono, U., & Suryanti, T. A. (2010). Indonesian adolescents' spiritual and religious experiences and their longitudinal relations with socioemotional functioning. *Developmental Psychology, 46*, 699-716. http://dx.doi.org/10.1037/a0018879

*Schneider, M., Dunton, G. F., & Cooper, D. M. (2008). Physical activity and physical self-concept among sedentary adolescent females: An intervention study. *Psychology of Sport and Exercise, 9*, 1-14. http://dx.doi.org/10.1016/j.psychsport.2007.01.003

Shanahan, M. J., Hill, P. L., Roberts, B. W., Eccles, J., & Friedman, H. S. (2014). Conscientiousness, health, and aging: The life course of personality model. *Developmental Psychology, 50*, 1407-1425. http://dx.doi.org/10.1037/a0031130

*Shapka, J. D., & Keating, D. P. (2005). Structure and change in self-concept during adolescence. *Canadian Journal of Behavioural Science, 37*, 83-96. http://dx.doi.org/10.1037/h0087247

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46*, 407-441. http://dx.doi.org/10.2307/1170010

*Siffert, A., Schwarz, B., & Stutz, M. (2012). Marital conflict and early adolescents' self-evaluation: The role of parenting quality and early adolescents' appraisals. *Journal of Youth and Adolescence, 41*, 749-763. http://dx.doi.org/10.1007/s10964-011-9703-1

*Silverthorn, N., DuBois, D. L., & Crombie, G. (2005). Self-perceptions of ability and achievement across the high school transition: Investigation of a state-trait model. *Journal of Experimental Education, 73*, 191-218. http://dx.doi.org/10.3200/JEXE.73.3.191-218

Simmons, R. G., Blyth, D. A., Van Cleave, E. F., & Bush, D. M. (1979). Entry into early adolescence: The impact of school structure, puberty, and early dating on self-esteem. *American Sociological Review, 44*, 948-967. http://dx.doi.org/10.2307/2094719

*Slutzky, C. B., & Simpkins, S. D. (2009). The link between children's sport participation and self-esteem: Exploring the mediating role of sport self-concept. *Psychology of Sport and Exercise, 10*, 381-389. http://dx.doi.org/10.1016/j.psychsport.2008.09.006

Soto, C. J. (2016). The Little Six personality dimensions from early childhood to early adulthood: Mean-level age and gender differences in parents' reports. *Journal of Personality, 84*, 409-422. http://dx.doi.org/10.1111/jopy.12168

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology, 100*, 330-348. http://dx.doi.org/10.1037/a0021717

Soto, C. J., & Tackett, J. L. (2015). Personality traits in childhood and adolescence: Structure, development, and outcomes. *Current Directions in Psychological Science, 24*, 358-362. http://dx.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*, 213-240. http://dx.doi.org/10.1037/a0028931

Spilt, J. L., van Lier, P. A. C., Leflot, G., Onghena, P., & Colpin, H. (2014). Children's social self-concept and internalizing problems: The influence of peers and teachers. *Child Development, 85*, 1248-1256. http://dx.doi.org/10.1111/cdev.12181

*Spray, C. M., Warburton, V. E., & Stebbings, J. (2013). Change in physical self-perceptions across the transition to secondary school: Relationships with perceived teacher-emphasised achievement goals in physical education. *Psychology of Sport and Exercise, 14*, 662-669. http://dx.doi.org/10.1016/j.psychsport.2013.05.001

*Steiger, A. E., Allemand, M., Robins, R. W., & Fend, H. A. (2014). Low and decreasing self-esteem during adolescence predict adult depression two decades later. *Journal of Personality and Social Psychology, 106*, 325-338. http://dx.doi.org/10.1037/a0035133

Sutton, A. J. (2009). Publication bias. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 435-452). New York, NY: Russell Sage Foundation.

Terracciano, A., McCrae, R. R., Brant, L. J., & Costa, P. T. (2005). Hierarchical linear modeling analyses of the NEO-PI-R scales in the Baltimore Longitudinal Study of Aging. *Psychology and Aging, 20*, 493-506. http://dx.doi.org/10.1037/0882-7974.20.3.493

Trzesniewski, K. H., & Donnellan, M. B. (2009). Reevaluating the evidence for increasingly positive self-views among high school students: More evidence for consistency across generations (1976–2006). *Psychological Science, 20*, 920-922. http://dx.doi.org/10.1111/j.1467-9280.2009.02361.x

Trzesniewski, K. H., & Donnellan, M. B. (2010). Rethinking “Generation Me”: A study of cohort effects from 1976-2006. *Perspectives on Psychological Science, 5*, 58-75. http://dx.doi.org/10.1177/1745691609356789

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*, 205-220. http://dx.doi.org/10.1037/0022-3514.84.1.205

Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2008). Do today's young people really think they are so extraordinary? An examination of secular trends in narcissism and self-enhancement. *Psychological Science, 19*, 181-188. http://dx.doi.org/10.1111/j.1467-9280.2008.02065.x

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.

Twenge, J. M., & Campbell, W. K. (2001). Age and birth cohort differences in self-esteem: A cross-temporal meta-analysis. *Personality and Social Psychology Review, 5*, 321-344. http://dx.doi.org/10.1207/S15327957PSPR0504_3

Twenge, J. M., & Campbell, W. K. (2008). Increases in positive self-views among high school students: Birth cohort changes in anticipated performance, self-satisfaction, self-liking, and self-competence. *Psychological Science, 19*, 1082-1086. http://dx.doi.org/10.1111/j.1467-9280.2008.02204.x

Twenge, J. M., & Campbell, W. K. (2010). Birth cohort differences in the Monitoring the Future Dataset and elsewhere: Further evidence for Generation Me—Commentary on Trzesniewski & Donnellan (2010). *Perspectives on Psychological Science, 5*, 81-88. http://dx.doi.org/10.1177/1745691609357015

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*, e9-e17. http://dx.doi.org/10.1037/pspp0000122

Twenge, J. M., Konrath, S., Foster, J. D., Campbell, W. K., & Bushman, B. J. (2008). Egos inflating over time: A cross-temporal meta-analysis of the Narcissistic Personality Inventory. *Journal of Personality, 76*, 875-901. http://dx.doi.org/10.1111/j.1467-6494.2008.00507.x

*Udell, W., Sandfort, T., Reitz, E., Bos, H., & Dekovic, M. (2010). The relationship between early sexual debut and psychosocial outcomes: A longitudinal study of Dutch

adolescents. *Archives of Sexual Behavior, 39*, 1133-1145. http://dx.doi.org/10.1007/s10508-009-9590-7

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

*Valkenburg, P. M., Koutamanis, M., & Vossen, H. G. M. (2017). The concurrent and longitudinal relationships between adolescents' use of social network sites and their social self-esteem. *Computers in Human Behavior, 76*, 35-41. http://dx.doi.org/10.1016/j.chb.2017.07.008

Van Dijk, M. P. A., Branje, S., Keijsers, L., Hawk, S. T., Hale, W. W., & Meeus, W. (2014). Self-concept clarity across adolescence: Longitudinal associations with open communication with parents and internalizing symptoms. *Journal of Youth and Adolescence, 43*, 1861-1876. http://dx.doi.org/10.1007/s10964-013-0055-x

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics, 30*, 261-293. http://dx.doi.org/10.3102/10769986030003261

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1-48. http://dx.doi.org/10.18637/jss.v036.i03

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1*, 112-125. http://dx.doi.org/10.1002/jrsm.11

*Viljaranta, J., Tolvanen, A., Aunola, K., & Nurmi, J. E. (2014). The developmental dynamics between interest, self-concept of ability, and academic performance. *Scandinavian*

Journal of Educational Research, 58, 734-756. http://dx.doi.org/10.1080/00313831.2014.904419

*von Soest, T., Wichstrom, 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*, 592-608. http://dx.doi.org/10.1037/pspp0000060

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*, 148-163. http://dx.doi.org/10.1037/a0029680

*Wichstrom, L., & von Soest, T. (2016). Reciprocal relations between body satisfaction and self-esteem: A large 13-year prospective study of adolescents. *Journal of Adolescence, 47*, 16-27. http://dx.doi.org/10.1016/j.adolescence.2015.12.003

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*, 10-18. http://dx.doi.org/10.1111/j.1750-8606.2009.00110.x

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*, 552-565. http://dx.doi.org/10.1037/0012-1649.27.4.552

Wilgenbusch, T., & Merrell, K. W. (1999). Gender differences in self-concept among children and adolescents: A meta-analysis of multidimensional studies. *School Psychology Quarterly, 14*, 101-120. http://dx.doi.org/10.1037/h0089000

*Wu, J., Watkins, D., & Hattie, J. (2010). Self-concept clarity: A longitudinal study of Hong Kong adolescents. *Personality and Individual Differences, 48*, 277-282. http://dx.doi.org/10.1016/j.paid.2009.10.011

*Young, J. F., & Mroczek, D. K. (2003). Predicting intraindividual self-concept trajectories during adolescence. *Journal of Adolescence, 26*, 586-600. http://dx.doi.org/10.1016/S0140-1971(03)00058-7

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. http://dx.doi.org/10.1016/j.jrp.2016.07.007

Footnotes

[1] At the time of coding, the qualifications of the coders were as follows: The second, fourth, and fifth author had a Master's degree in psychology and the third author had a Ph.D. degree in psychology.

[2] When coding the first set of studies included in this meta-analysis (1806 to September 2015), interrater agreement was lower ($\kappa = .73$) for the variable sample type. As reported in Orth et al. (2018), the disagreement resulted from overlap between two categories, specifically "community samples (convenience)" and "community samples (regionally representative);" regionally representative was defined as representative for a region such as a county or city. Given the overlap, we merged these two categories into one category (denoted as community sample), resulting in high agreement for the revised variable of sample type ($\kappa = 1.00$). When coding the second set of studies included in this meta-analysis (September 2015 to October 2018), we used the revised variable, distinguishing the following categories: nationally representative, community, and college students.

[3] In the analyses, we did not consider data from one sample from middle adulthood. More precisely, in this sample the mean age was 49.85 years at Time 1, and the study included a measure of physical appearance (Guérin, Goldfield, & Prud'homme, 2017). We ignored the data from this sample because its age differed fundamentally from the age of all other eligible samples, which ranged from 4.9 to 26.0 years as reported above. Inclusion of a single study from middle adulthood would not have allowed for any reliable conclusions about mean-level change in this developmental period. We did not examine the data from this sample and did not include its characteristics in the description of studies.

[4] We used this contrast because the number of samples was large for the United States, whereas the number of samples was relatively small for all other countries, corresponding to the procedure used in the meta-analysis on the development of global self-esteem reported in Orth et al. (2018). We did not test other contrasts for country to avoid an increased risk of false-positive findings.