



Disentangling the Long-Term Compositional Effects of School-Average Achievement and SES: a Substantive-Methodological Synergy

Herbert W. Marsh¹ · Reinhard Pekrun^{1,2,3} · Theresa Dicke¹ · Jiesi Guo¹ · Philip D. Parker¹ · Geetanjali Basarkod¹

Accepted: 10 October 2022
© The Author(s) 2023

Abstract

We juxtapose (positive and negative) compositional effects of school-average achievement and school-average socioeconomic status (SES) on students' academic self-concept (ASC), final high-school grade-point-average (GPA), and long-term outcomes at age 26 (educational attainment and educational and occupational expectations). We used doubly-latent multilevel compositional models with a large, nationally representative longitudinal sample (16,197 Year-10 students from 751 US high schools), controlling background variables (gender, age, ethnicity, academic track, and a composite risk factor). At the individual-student level, the effects of achievement, SES, ASC, and GPA on long-term outcomes were consistently positive. However, mostly consistent with a priori theoretical predictions, (1) the compositional effects of school-average achievement on ASC, GPA, and educational and occupational expectations were significantly negative (although non-significant for final attainment); (2) the compositional effects of school-average SES on ASC, educational attainment, and educational and occupational expectations were significantly positive (but nonsignificant for GPA); and (3) the compositional effects on long-term outcomes were partly mediated by ASC and particularly by GPA. These findings demonstrate that the positive effects of school-average SES are distinguishable from the adverse effects of school-average achievement. We discuss how these findings extend Göllner et al.'s (Psychological Science 29:1785–1796, 2018) highly controversial conclusion regarding the benefits of schools with high school-average SES but low school-average achievement. We also relate our research to Luthar et al.'s (American Psychologist 75:983–995, 2020) findings of adverse mental health problems associated with attending high-achieving schools. Our results have important implications not only for theory and methodology but also for parents' selection of schools for their children and policy regarding the structure of schools (a substantive-methodological synergy).

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

Extended author information available on the last page of the article

Keywords School-average achievement · SES · Big-fish-little-pond effect · Long-term educational and occupational expectations · Long-term educational attainment · Substantive-methodological synergy

Conventional wisdom suggests academic benefits in attending selective schools, high-achieving schools (i.e., schools with high school-average achievement), and high-SES schools (i.e., schools with high school-average SES). The validity of this assumption has important policy implications for how schools are structured, for school funding, and for parents' choices about the schools their children attend. However, a growing body of research calls this assumption into question. We will review and extend this research, further challenging the validity of conventional wisdom. Furthermore, there is a lack of clarity as to whether the critical school compositional variable is school-average achievement or school-average socioeconomic status (SES). We aim to clarify this question.

More specifically, our study is aimed at evaluating *a priori* predictions on the short- and long-term compositional effects of school-average achievement and SES. Our focus is not on the effects of individual-student achievement and SES. However, school compositional effects are defined as the effects of school-average variables on outcomes beyond the contributions of individual-student characteristics to explaining these outcomes (e.g., Becker et al., 2022; Harker & Tymms, 2004). Hence, evaluating the effects of individual-student achievement and SES is also important.

As an advanced organizer, we briefly summarize the variables to be considered and the conceptual model that guides our literature review and subsequent analyses (Fig. 1; also see Supplemental Materials, Sect. 1, and Appendix). Outcome variables include academic self-concept (ASC), final high school grade point average (GPA), and long-term outcomes at age 26 (educational attainment; educational and occupational expectations; Fig. 1). In analyzing compositional effects on these outcomes, we control individual-level student achievement, family SES, demographic control variables (gender, age, academic track, and ethnicity), and a composite risk factor. We tested the compositional effects with a large nationally representative sample of US students (16,197) and schools (751).

Our review of the literature summarized below identifies serious methodological issues in how many studies evaluated compositional effects. School-compositional effects are inherently a multilevel issue (i.e., students nested within schools) and must be assessed with appropriate multilevel models. The methodologically strongest school-compositional research employs latent multilevel models (e.g., Becker et al., 2022; Lüdtke et al., 2008, 2011). The most consistent school compositional findings from multilevel analyses are the adverse effects of school-average achievement on ASC—the big-fish-little-pond-effect (BFLPE; Marsh & Seaton, 2015; Marsh et al., 2021a, b). There is a robust theoretical basis for the BFLPE that is generally useful in evaluating school-compositional effects. Hence, we begin with a brief review of the relevance of the BFLPE, its theoretical model, and related

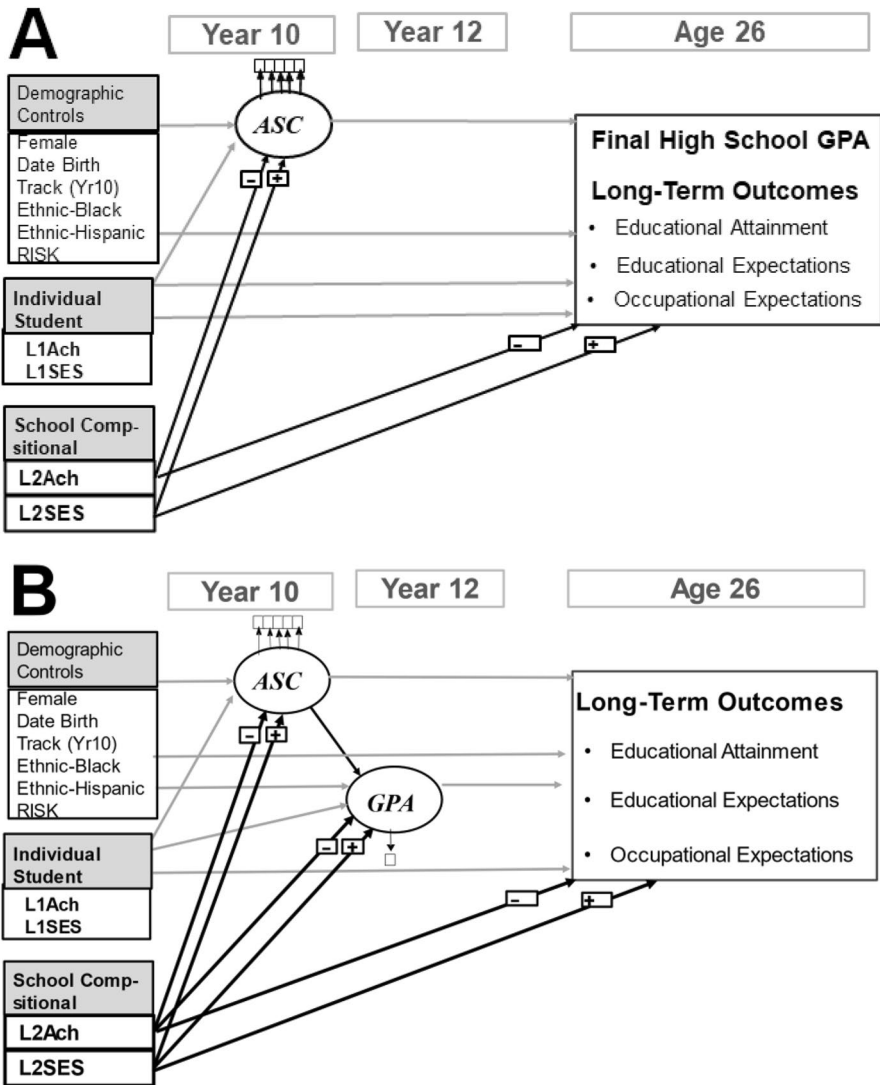


Fig. 1 Schematic diagrams of variables and their temporal ordering considered in the present investigation. *Note.* L1: individual-student level; L2: school-average level; SES: socioeconomic status (at L1 and L2 levels); Ach: achievement (at L1 and L2 levels); GPA: grade point average; Trk: academic track. **A** is the conceptual basis of Models M3-M5 (Table 1). **B** is the conceptual basis of Model M6 (Tables 3 and 4) that evaluates the indirect effects of achievement and SES on long-term outcomes mediated through ASC and GPA

methodology. We then apply lessons from this research to the broader issue of the long-term effects of school-average achievement and SES on educational attainment, educational expectations, and occupational expectations (Fig. 1).

Literature on Negative Effects of School-Average Achievement on ASC: A Prototype Compositional Effect

The Relevance of Academic Self-Concept to Academic Outcomes

ASCs are a student's self-perceptions of academic competence and accomplishments. Positive ASCs and the need to feel a sense of competence are as follows: "a basic psychological need that has a pervasive impact on daily life, cognition and behavior, across age and culture...an ideal cornerstone on which to rest the achievement motivation literature but also a foundational building block for any theory of personality, development, and well-being" (Elliot & Dweck, 2005, p. 8). Self-concept is a "cornerstone of both social and emotional development" in early childhood (Kagan et al., 1995, p. 18), "a major (perhaps the major) structure of personality" (Greenwald, 1988, p. 30), and a driving force for the positive psychology movement (Marsh & Craven, 2006).

As such, ASCs are an important educational outcome. However, they also contribute to the prediction of long-term educational attainment, beyond the effects of SES, IQ, school grades, and standardized achievement tests (see Marsh, 2007; Marsh et al., 2020; Marsh & Seaton, 2015). For example, Wouters et al. (2011) showed that ASC in high school affected success and adjustment in higher education beyond the effects of high-school achievement and control variables. In addition, longitudinal cross-lagged panel models show that ASC and achievement are reciprocally related over time; each is a cause and an effect of the other (see meta-analyses by Huang, 2011; Valentine et al., 2004).

Gutman and Schoon (2013) argued that noncognitive skills—including positive self-beliefs and ASC—are as important, or even more important, than cognitive skills in explaining academic and employment outcomes (also see Heckman et al., 2006; Heckman & Rubinstein, 2001). Heckman et al. (2006) further argued that early intervention programs' success is due to their impact on noncognitive variables rather than cognitive skills. Furthermore, across 26 countries and 14 noncognitive factors (self-regulated learning strategies, self-beliefs, motivation, and learning preferences), achievement correlated most highly with ASC (Marsh, 2007; Marsh & Craven, 2006). Thus, among various noncognitive skills, ASCs and positive self-beliefs are especially important for explaining achievement.

The Big-Fish-Little-Pond Effect (BFLPE): A Prototypical School Compositional Effect

ASC relates positively to academic achievement and predicts short-term and long-term academic outcomes. Nevertheless, contrary to the expectations of many parents, students, teachers, policymakers, and even some educational researchers, the effects of attending academically selective schools and classes on ASC are adverse

(BFLPE). Moreover, although the impact of individual-student achievement on ASC is positive, the effect of school-average achievement is negative.

Since the early BFLPE studies in the 1980s, there has been a wealth of support for BFLPE predictions based on studies that used different experimental and analytical approaches (Alicke et al., 2010; Fang et al., 2018; Marsh & Seaton, 2015; Marsh et al., 2008; Marsh et al., 2012b; Zell & Alicke, 2010). Marsh et al. (2020) reviewed findings based on four Programme for International Student Assessment (PISA) data collections with 1.25 million students. The results show that the effect of school-average achievement on ASC was negative in all but one of the 191 samples representing different countries/regions and significantly so in 181 samples. The BFLPE tends to increase in size during high school (Marsh et al., 2001). Furthermore, in two studies, Marsh (2007) showed that the BFLPE formed in high school is as large or even larger two and four years after graduation from high school. Frank (1985, 2012) provides an evolutionary argument for the universality of social comparison processes underpinning the BFLPE (Marsh et al., 2021a, b; also see Festinger, 1954). This research literature demonstrates that the BFLPE is one of educational and psychological research's most robust and consistent findings (e.g., Fang et al., 2018; Marsh & Seaton, 2015; Marsh et al., 2021a, b) and an ideal foundation for building school-compositional studies more broadly.

Methodological Basis: Doubly-Latent Multilevel Models

Methodologically, researchers seized upon the BFLPE as a classic application of multilevel analysis—the same variable having opposite effects at the individual and school- or class-average levels. BFLPE research demonstrates why separating the effects of individual-student achievement and group-average achievement is vital. Furthermore, ongoing BFLPE research has contributed significantly to developing sophisticated and more appropriate statistical models.

School compositional effects are appropriately evaluated with doubly-latent multilevel structural equation models (SEMs) that are latent for individual-student and school-average outcomes and control preexisting differences. In this respect, these models control measurement error and preexisting differences that are potential biases in estimating the effects of school-average achievement (Lüdtke et al., 2008, 2011; Marsh et al., 2009, 2012). Doubly-latent SEMs have important implications for evaluating compositional effects on many outcomes. The doubly-latent multilevel SEMs based on BFLPE research have led to the current best practice in evaluating compositional effects (e.g., Becker et al., 2022; Lüdtke et al., 2008, 2011). Here, we build on this research to distinguish between compositional effects based on school-average achievement and school-average SES.

Theoretical Basis: Social Comparison Processes

Since James (1890/1983), psychologists have recognized that individuals evaluate objective accomplishments compared to frames of reference. Thus, James indicated,

“we have the paradox of a man shamed to death because he is only the second pugilist or the second oarsman in the world” (1890/1963, p. 310). Marsh proposed the BFLPE to encapsulate frame-of-reference effects (Marsh & Parker, 1984; Marsh et al., 2008). He based this on an integration of theoretical models and empirical research from diverse disciplines: relative deprivation theory (Davis, 1966; Stouffer et al., 1949), sociology (Alwin & Otto, 1977; Hyman, 1942), psychophysical judgment (e.g., Helson, 1964; Parducci, 1995), social judgment (e.g., Morse & Gergen, 1970; Sherif & Sherif, 1969; Upshaw, 1969), and social comparison theory (Festinger, 1954).

The BFLPE model (Marsh & Seaton, 2015) hypothesizes that students compare their own achievements with the achievements of their classmates and use this social comparison impression as one basis for forming their own ASC (Fig. 1). Individual achievement positively predicts ASC (the better I perform, the higher my ASC). In contrast, school-average achievement negatively predicts ASC (the brighter my classmates, the lower my ASC). Hence, ASC depends on a student’s own academic accomplishments and those of their classmates. According to the BFLPE, students who attend schools and classes with a high average achievement will have lower ASCs than equally able students attending mixed- or low-ability schools and classes. This implies an adverse effect of class/school-average achievement on ASC. Consistent with social comparison theory, the size of the BFLPE is determined substantially by the extent of ability stratification in schools (Parker et al., 2021). If all schools had the same school-average achievement, the BFLPE would disappear.

Competing Effects of Contrast and Assimilation

Social psychologists hypothesize contrast and assimilation as two competing forces associated with compositional effects (e.g., Diener & Fujita, 1997; Kelley, 1952; Suls & Wheeler, 2000). Contrast processes operate when people’s perceptions, opinions, or behavior depends on their perceived relative (rank) position within their group, particularly for self-evaluation variables (Kelley, 1952; Marsh et al., 2020; Parker et al., 2018). Contrast effects are the basis of the negative BFLPE. Assimilation processes operate when people form their perceptions, opinions, or behaviors according to group norms. Kelley (1952) suggested these processes are more likely to drive identity, values, and behavior variables, such that individuals become more like the group to which they belong.

Assimilation theories argue that attending selective schools will benefit students beyond what is explained by the—often substantial—preexisting advantages (e.g., high individual achievement and SES; see Göllner et al., 2018; Marsh, 1991, 2007). The potentially positive effects of selective schools might partly be due to the typically better resources in these schools. However, the so-called positive peer spillover (or peer contagion) effects attributed to selective schools (e.g., Harris, 2010; Mayer & Jencks, 1989) are particularly relevant to assimilation theories. According to this perspective, interacting with the more advantaged peers in selective schools and related social networks will rub off on all students, resulting in long-term benefits. Conversely, contrast theories argue that social comparison and frame-of-reference

effects associated with attending selective schools will adversely affect academic self-beliefs and long-term outcomes related to these self-beliefs (Göllner et al., 2018; Marsh, 1991, 2007).

Contrast and assimilation effects can operate simultaneously. For example, expanding the BFLPE model, Marsh (1987; Marsh et al., 2000) noted that being an average-ability student in a high-ability group of classmates may affect ASC such that it is (a) below average because the frame-of-reference is established by the performance of above-average students (i.e., a contrast effect, the BFLPE effect); (b) above average as a consequence of membership in the high-ability group (i.e., an assimilation effect, a reflected glory or group identification effect); (c) average because it is unaffected by the immediate context of the other students; or (d) average because (a) and (b) both occur and cancel each other. In this respect, the negative BFLPE actually observed could be the net effect of a large negative (contrast) effect and a smaller positive (assimilation) effect.

Literature on Broadening the Perspective: Multiple Outcomes and Multiple Compositional Effects

Effects of School-Average Achievement on Outcomes Beyond ASC

The BFLPE is highly robust (Fang et al., 2018; Marsh et al., 2020, 2021a, b; Marsh & Seaton, 2015), but specific to the adverse effects of school-average achievement on ASC and related academic self-beliefs. Hence, a critical question is how school-average achievement affects other outcomes, such as individual achievement and postsecondary outcomes. Thus, Marsh (1991) evaluated school-average achievement effects on a wide array of outcomes in a large, nationally representative, longitudinal study of US high school students. Students were surveyed in Year 10, Year 12, and again two years after graduation from high school. After controlling background variables and initial achievement, the effects of school-average achievement were negative for almost all Year 10, Year 12, and postsecondary outcomes: 15 of the 17 effects were significantly negative, and only two were nonsignificant. School-average achievement most negatively affected ASC (the BFLPE) and educational aspirations, but also negatively affected general self-concept, advanced coursework selection, school grades, academic effort, standardized test scores, occupational aspirations, and subsequent actual college attendance two years after high school graduation. In each case, these adverse effects were partially explicable by diminished ASCs. These results suggest that the adverse effects of attending academically selective schools extend well beyond those for ASC. In related research, Espenshade et al. (2005) found that entrance into elite US universities was positively associated with individual-student achievement but negatively related to school-average levels of achievement. The school's reputation had a counterbalancing assimilation-like effect, but this effect was small.

In related research, Luthar et al. (2020) argued that students in high-achieving schools are an “at-risk group” based on converging evidence on social comparison

processes and two major national policy reports. Complementing the focus of BFLPE research on academic outcomes, Luthar et al. emphasized the negative effects of high-achieving schools on nonacademic outcomes (e.g., mental health, psychological problems, and psychological well-being). Relatedly, Pekrun et al. (2019) evaluated the effects of school-average achievement on students' academic emotions. Three studies found that individual-student achievement related positively to positive emotions (enjoyment, pride) and negatively to negative emotions (anger, anxiety, shame, and hopelessness), thus showing beneficial effects. In contrast, class-level achievement adversely impacted both positive and negative emotions. Pekrun et al., (2019, p. 166) concluded that: "individual success drives emotional well-being, whereas placing individuals in high-achieving groups can undermine well-being. Thus, the findings challenge policy and practice decisions on the achievement-contingent allocation of individuals to groups."

Effects of School-Average Achievement on Subsequent Achievement

Based on his extensive meta-analytic research, Hattie (2002) reported that tracking (i.e., grouping according to ability) has almost no effect on subsequent achievement. He argued that any small positive compositional effects of attending high-track schools are likely to result from uncontrolled variables (e.g., preexisting differences between students and differences in resources and curriculum). In contrast, he emphasized that the adverse effects of school-average achievement on ASC (the BFLPE) were particularly robust.

The doubly-latent multilevel SEM routinely applied in BFLPE studies has important implications for testing school-compositional effects on achievement. For example, in an early study of the impact of school-average achievement, Harker and Tymms (2004) found that apparently positive school-average achievement effects disappeared with appropriate control for measurement error and covariates. They referred to positive school-average achievement effects as "phantom effects"—now you see them, now you don't. Here, we use the term phantom effects to represent the positive bias in apparently positive effects of school-average achievement that are actually due to the failure to control for measurement error and preexisting differences. In estimating the effects of school-average achievement, these phantom effects inevitably will be positive. Furthermore, observational studies will always have at least some residual phantom effects. Critically, if phantom effects are sufficiently large, controlling them can shift positively biased estimates for effects of school-average achievement on subsequent individual achievement from positive to nonsignificant or even negative.

The findings from several recent studies that used the doubly-latent model and controlled for covariates (including prior achievement) are consistent with this interpretation (e.g., Dicke et al., 2018; Televantou et al., 2015, 2021). In each study, controlling measurement error and covariates led to the following: (a) school-average achievement effects on ASC becoming more negative, and (b) school-average achievement effects on subsequent achievement becoming less positive, nonsignificant, or even negative. Thus, Dicke et al. (2018, p. 1112) found that: "More appropriate multilevel modeling that

controls for phantom effects (due to measurement error and pre-existing differences) makes the BFLPE even more negative, but turns the peer spillover effect from positive to slightly below zero. Thus, attending a high-achieving school negatively affects academic self-concept and has a nonpositive effect on achievement.” These studies question previous studies and meta-analyses that showed a positive peer spillover effect but did not control phantom effects.

Becker et al. (2022) presented results based on five large, nationally representative German datasets. Following the Dicke et al. (2018) recommendations, they used doubly-latent multilevel SEMs with covariates to control measurement error and bias associated with pre-existing differences (also referred to as selection bias). Across the five datasets, there were positive effects of school-average achievement on subsequent achievement. However, the estimates changed when controlling for academic track (primarily based on achievement in primary school before the start of secondary school in Germany). After controlling track, the effects of school-average achievement were minimal and only marginally significant (average effect size was 0.06; 95% CI=0.01 to 0.11). For a total of 15 outcomes across the five databases, only five were significantly positive, and one was significantly negative. Furthermore, Becker et al. noted that differences between findings across studies suggest that compositional effects of school-average achievement may vary. They called for more research to identify conditions that explain these differences.

School Selectivity: Juxtaposing the Effects of School-Average Achievement and SES

The effects of achievement and SES are substantially correlated and difficult to disentangle at both the individual-student and school-average levels. For example, in their review of sociological research on educational and occupational aspirations, Alwin and Otto (1977) reported adverse effects of school-average achievement but positive effects of school-average SES. Bachman and O'Malley (1986, p. 35) similarly emphasized the importance of disentangling the effects of school-average achievement and SES, noting that “two different types of school context effects on such outcome variables as college plans and occupational aspirations...The ability context of the school shows negative effects, but the school socioeconomic context shows positive effects (Alwin & Otto, 1977; Meyer, 1970).”

Marsh (1991) reviewed psychological research on school-average achievement effects and sociological research on school-average SES effects. He predicted and found that school-average achievement effects were consistently more negative than school-average SES effects across a broad range of educational outcomes. Indeed, consistent with Alwin and Otto's (1977) and Bachman and O'Malley's (1986) conclusions, Marsh (1991) showed that school-average SES positively affected ASC, coursework selection, test scores, educational and occupational aspirations, and subsequent university attendance. Moreover, he contrasted relatively larger adverse effects of school-average achievement and relatively smaller positive effects of school-average SES. These were consistent with Alwin and Otto's review and

previous results from the Youth in Transition study (Marsh, 1987; also see Marsh & O'Mara, 2010).

Nevertheless, Marsh (1991) argued that school-average achievement and school-average SES often are correlated so highly that the compositional effects of each are difficult to disentangle. Thus, for example, Sirin's (2005) subsequent meta-analysis reported that achievement and SES were only moderately correlated (mean $r=0.28$) at the individual-student level. However, school-average achievement and SES were substantially correlated (mean $r=0.67$). Marsh noted a need for more research, using potentially more robust statistical models, to disentangle the two compositional effects of these two school-average variables.

Marsh and O'Mara (2010) noted that few studies had included both achievement and SES at both levels (i.e., individual achievement and SES, and school-average achievement and SES) in the same model (also see Göllner et al., 2018). Critically, for disentangling individual-student level effects and school-average compositional effects, it is necessary to include all four variables. Previous research has not always done this. For example, Bachman and O'Malley (1986) included individual-student achievement and SES as well as school-average achievement, but not school-average SES. Marsh (1987) considered all four variables, but tested the effects of achievement and SES separately. Alwin and Otto (1977) considered school-average achievement and SES in the same model but not the corresponding student-level variables.

However, Marsh (1991) did include all compositional variables. He found that school-average SES and individual-student achievement and SES generally exhibited positive effects. However, the effects of school-average achievement were typically negative, across a range of educational outcomes. Marsh and O'Mara (2010) reported similar results in their reanalysis of the Youth in Transition study. Their review and results suggested that students identify with higher levels of school-average SES (assimilation or reflected glory effect) but contrast themselves with higher levels of school-average achievement (contrast effect). Marsh and O'Mara concluded that "this juxtaposition between school-average SES, school-average ability, assimilation, and contrast is an important topic for further research" (p. 65). However, none of these early studies used doubly-latent multilevel SEMs controlling measurement error and appropriate covariates.

Returning to this classic issue, Göllner et al. (2018) emphasized that conventional wisdom suggests that attending high-SES schools contributes to students' long-term success (e.g., Coleman et al., 1966; Coleman & Hoffer, 1987). Göllner et al. discussed possible advantages of "good schools" regarding school facilities, including better teachers, but also contagion effects (assimilation; positive peer spillover effects). However, Göllner et al. also lamented that compositional studies of school-average achievement rarely considered school-average SES, and studies of school-average SES rarely considered school-average achievement. Their own analysis used archive data from Project TALENT. The data include test scores and educational expectations collected in 1960 when students were in grades 9–12 (mean year in school = 10.4, $SD=1.11$). Postsecondary outcomes were from the 11-year follow-up (response rate 20%) and the 50-year follow-up (1% response rate). Göllner et al. included individual achievement

and SES as well as school-average achievement and SES in the analysis, controlling for three demographic variables (year in school, gender, and ethnicity). They used full-information maximum-likelihood estimation based on all variables in their model to control for the substantial amount of missing data.

Consistent with earlier studies, Göllner et al. (2018) found school-average SES positively affected educational expectations, attainment, and occupational status. In contrast, school-average achievement had largely negative effects on these outcomes. The unique contribution of this study is that the positive effects of school-average SES were evident even in the 50-year follow-up. Göllner et al. suggested that these effects reflected the positive impact of learning resources as well as positive peer spillover effects (assimilation effects). Conversely, the negative effects of school-average achievement reflected the adverse impact of social comparison processes (i.e., contrast effects like those that are the basis of the BFLPE). Göllner et al., (2018, p. 10) concluded: “it appears that the optimal combination would be a school with a high socioeconomic composition combined with a modest achievement composition” and “Students who attend more socioeconomically advantaged schools benefit from the positive social environment but can be harmed if a high socioeconomic composition is combined with a high achievement composition.” Given the highly controversial nature of their conclusion, they noted caution and the need for further research. As reasons for caution, they highlighted their use of historical data (students in 1960), inherent difficulties of the Project TALENT data, and complications in disentangling school-average achievement from school-average SES due to their very high correlation (also see von Keyserlingk et al., 2020).

The Present Investigation: Research Hypotheses

Our overarching aim is to disentangle the short- and long-term effects of school-average academic achievement (L2Ach) and school-average SES (L2SES), controlling individual-student achievement (L1Ach), individual-student SES (L1SES), and demographic variables. We use the longitudinal data from the US Educational Longitudinal Survey 2002 (ELS:2002). The sample consisted of high school students first assessed in Year 10 and followed up through age 26 (see Fig. 1). Achievement and SES were measured in Year 10. Outcome variables (Fig. 1) are ASC and GPA in Year 10 and long-term educational attainment, educational expectations, and occupational expectations assessed at age 26. Although the effects of L1Ach, L1SES, and the covariates on these outcomes are important in testing our model, our primary focus is on the compositional effects of school-average achievement and SES. For these effects, we offer the following hypotheses.

- H1: school-average achievement negatively predicts all outcomes (ASC, GPA, and age-26 outcomes; Fig. 1A).
- H2: school-average SES positively predicts all outcomes (ASC, GPA, and age-26 outcomes; Fig. 1A).

- H3: the effects of student-level and school-average achievement and SES on long-term outcomes at age 26 are mediated in part through ASC and GPA (Fig. 1B).

Method

Sample

We used the public US ELS:2002 database ($N=16,197$ high school Year-10 students from 751 schools followed up through age 26; see Ingels et al., 2004, 2005, 2007, 2014). Recruitment of students was based on a nationally representative probability sample of public, Catholic, and other private schools in the spring term of the 2001–02 school year. ELS:2002 employed a two-stage complex sample design. They first selected schools and then selected Year-10 students (mostly 15-year-olds) within each school. For further discussion of the sample and variables, see Supplemental Materials, Sect. 2; also see the ELS:2002 website for study design, variables, and studies using these data (<https://nces.ed.gov/surveys/els2002/>).

In late 2004 and 2005 (i.e., one year after most students had graduated from high school), ELS:2002 had a 91% response rate when requesting official school transcripts. In 2012, when most participants were 26, data collection focused on actual educational attainment at this point and on participants' future educational and occupational expectations of their career status at age 30. Through concerted data collection activities and procedures (Ingels et al., 2014), ELS:2002 achieved a response rate of 78.2% in the 2012 data collection. Furthermore, ELS:2002 supplemented information about cohort members from extant data sources such as the American Council on Education and the U.S. Department of Education Central Processing System.

Measures

Compositional Predictor Variables

We used Year 10 achievement and SES at the individual-student level (L1) and the school level (L2) as predictors to estimate compositional effects (see Fig. 1). ELS:2002's measure of SES is a composite index based on five standardized scores: father's/guardian's education, mother's/guardian's education, father's/guardian's occupation, mother's/guardian's occupation, and family income. ELS:2002 used parent data when available and student data if parent data were missing. In some cases, ELS:2002 imputed data from other materials. We aggregated the scores for individual SES (L1SES) to the school level to form L2SES. We used ELS:2002's standardized test measures to represent math and reading achievement. We constrained reading and

math to be equally weighted in constructing L1Ach scores in our statistical models. The L1Ach scores were aggregated within schools to form L2Ach.

Outcome Variables

The outcome variables are ASC in Year 10, GPA at the end of high school, educational attainment at age 26, and long-term educational and occupational expectations at age 26 (see Fig. 1). We assessed ASC with the following five ELS:2002 items: *When I sit myself down to learn something really hard, I can learn it; If I decide not to get any bad grades, I can really do it; If I want to learn something well, I can; When I study, I make sure that I remember the most important things; When studying, I try to do my best to acquire the knowledge and skills taught.* Participants responded to each item using a 4-point scale ranging from 1 (*almost never*) to 4 (*almost always*). Higher scores reflect more favorable ASCs.

ELS:2002 requested that schools provide academic transcripts for all participating students. They used these transcripts to compute a final GPA that was comparable across schools.

ELS:2002 assessed educational attainment and educational and occupational expectations at age 26 in the follow-up questionnaire. The assessment was either a self-administered web-based survey or a computer-assisted interview. Although the survey was the primary source of information, ELS:2002 used other sources of information when the survey data was unavailable to check the consistency of survey responses (Ingels et al., 2014). Final educational attainment at age-26 was coded according to the following 9-category response scale: 1=no high school credential or postsecondary attendance; 2=high school credential, no postsecondary attendance; 3=some postsecondary attendance but no postsecondary credential; 4=undergraduate certificate or diploma; 5=associates degree; 6=bachelor's degree; 7=postbaccalaureate certificate; 8=master's degree/postmaster's certificate; and 9=doctoral degree. Respondents reported the highest level of education they expected to achieve by age 30 and their expected occupation at age 30. ELS:2002 coded educational expectations with a 7-category response scale: less than high school graduation, high school diploma or General Educational Development equivalent, undergraduate certificate or diploma, associates degree, bachelor's degree, master's degree, and doctoral degree. ELS:2002 coded occupational expectations according to occupational prestige.

Demographic Control Variables

In his methodologically oriented review of the best practice concerning the inclusion of covariates, VanderWeele (2019) noted that a broad range of demographic variables should be included. This inclusive strategy of using demographic control variables is consistent with recommendations that Lütke and Robitzsch (2021) derived from their methodological analysis of longitudinal panel study designs. There is substantive interest in how these demographic control variables (particularly gender)

relate to our study variables. However, our primary focus is to use these demographic variables to control for preexisting differences and to evaluate how their inclusion affects estimated compositional effects.

For present purposes, demographic control variables consisted of gender, age, track in Year 10 (1 = academic track; 0 = nonacademic track), two dichotomous variables representing ethnicity (Black, 1 = yes, 0 = no; Hispanic, 1 = yes, 0 = no), and a composite risk factor compiled by ELS:2002. The risk factor consists of six indicators: (1) comes from a single-parent household, (2) has two parents without a high school diploma, (3) has a sibling who has dropped out of school, (4) has changed schools two or more times (excluding changes due to school promotions), (5) has repeated at least one grade, and (6) comes from a household with an income below the federal threshold for poverty. In some cases, the scores making up the risk factor were imputed by ELS:2002 using data not available in the public ELS:2002 database. To avoid confusion, we use the term control variables when referring to this set of background variables and refer separately to individual SES and achievement that we also controlled in estimating compositional effects.

Statistical Analyses

We used multilevel (SEMs) to estimate compositional effects using Mplus (Version 8.4; Muthén & Muthén, 2017). We estimated doubly-latent two-level random-intercept models (L1: students; L2: schools) based on the framework proposed by Lüdtke and colleagues (Lüdtke et al., 2008, 2011; Marsh et al., 2009, 2012a, b). We used the robust maximum likelihood estimator (MLR). This estimator is robust against any violations of normality assumptions and uses weights to adjust for unequal probabilities of student selection. To facilitate the interpretation of the parameter estimates, we standardized all continuous variables across the student sample ($M=0$, $SD=1$). In addition, we scaled latent factors so that the variance of each factor was approximately 1.0. This resulted in parameter estimates that were scaled relative to a common metric and represented standardized effects that facilitated interpretations.

As is typical in large-scale longitudinal field studies, a substantial portion of the sample had some missing data. Across all variables considered here, coverage rates varied from 66 to 100% (see Supplemental Table 1). However, we did not exclude any cases because of missing data but used multiple imputation. Multiple imputation results in trustworthy, unbiased estimates for missing values, even in the case of large numbers of missing data (Enders, 2010). It is an appropriate method to manage missing data in large-scale longitudinal studies (Jelčić et al., 2009). More specifically, under the missing-at-random (MAR) assumption, missingness is allowed to be conditional on all variables included in the analysis (e.g., Newman, 2014). In other words, the critical situation of not-MAR is when missingness is dependent on the variable for which data are missing. For longitudinal data, this implies that missing values are allowed to be conditional on the same variable's values collected in a different wave. This feature of longitudinal data makes it unlikely that MAR assumptions are seriously violated. An important advantage of the multiple imputation approach to

Table 1 Goodness of fit for alternative models

Model	Model description					Model fit				
	L1SES	L2SES	L1Ach	L2Ach	6 Co-vars	Chi-Sq	df	RMSEA	CFI	TLI
L1 measurement models										
M1 L1 CFA 5 ASC items	Yes	No	Yes	No	No	651	5	.089	.977	.955
M2 L1 CFA all variables	Yes	No	Yes	No	Yes	2025	79	.039	.979	.948
ACH as a predictor at L1 & L2										
M3 BFLPE + L1SES + 6Covars	Yes	No	Yes	Yes	Yes	2390	107	.036	.975	.951
SES as a predictor at L1 & L2										
M4 L2SES: L1Ach + 6Covars	Yes	Yes	Yes	No	Yes	2365	98	.038	.975	.950
Ach + SES as predictors at L1 & L2										
M5 L2Ach + L2SES + 6Covars	Yes	Yes	Yes	Yes	Yes	2400	112	.036	.974	.949
Indirect effect models										
M6 L2ACH + L2SES + 6COVs	Yes	Yes	Yes	Yes	Yes	2400	112	.036	.974	.949

Summary of goodness-of-fit statistics for the models (M1–M6) considered in the present investigation

L1 individual-student level, *L2* school level, *Ach* achievement (standardized test scores), *SES* socioeconomic status, *ASC* academic self-concept, *BLFPE* big-fish-little-pond effect, *Covar* covariate, *CFA* confirmatory factor analysis, *Chi* chi-square, *df* degrees of freedom, *RMSEA* root mean square error of approximation, *CFI* comparative fit index, *TLI* Tucker-Lewis index

In Supplemental Materials (Sect. 2), we present additional models, supplementing Models M3–M6 without controlling the set of 6 demographic control variables, and the inclusion of L1 and L2 measures of SES and achievement

missing data is that the control for missingness is used consistently across models based on different variables. Here, we used the Mplus two-level imputation procedure supplemented by auxiliary variables to create 20 imputed datasets (Asparouhov & Muthen, 2010).

We estimated 6 models (see Table 1 and Fig. 1; also see supplemental models in Supplemental Materials, Sect. 6). In two measurement models (M1 and M2), we used confirmatory factor analysis to test the factor structure of the ASC scale. Model M1 included only this scale. In Model M2, we add all the other study variables (single-item variables). Model M3 estimated the compositional effects of L2ACH, including both L1SES and the covariates. Thus, Model 3 fully controlled the effects of the pre-existing differences between students assessed in the project. Following the same logic, Model M4 estimated the compositional effects of L2SES, controlling L1 ACH and the covariates.

Model M5 included L2ACH and L2SES, thus making it possible to compare their unique compositional effects to the effects estimated in Models M3 and M4, which only considered one of the two variables. Finally, Model M6 tested the mediation of long-term effects (Hypothesis 3). ASC and GPA mediated the effects of L2ACH and L2SES on age-26 variables (see Fig. 1B).

We evaluated model fit with fit indices that are relatively sample-size independent (Hu & Bentler, 1999; Marsh et al., 2004), including the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the Tucker-Lewis index (TLI). Values smaller than 0.08 and 0.06 for the RMSEA support acceptable and good model fits, respectively. Population values of TLI and CFI vary along a 0–1 continuum, in which values greater than 0.90 and 0.95 typically reflect good and excellent fits to the data, respectively. Nevertheless, these recommended cut-off values constitute only rough descriptive guidelines rather than “golden rules” (Marsh et al., 2004).

Preliminary Analyses: Fit of the Structural Equation Models

Measurement Models (M1 and M2)

We estimated the two measurement models (M1 and M2 in Table 1) at the individual student level. We tested single-level models for these preliminary analyses using the Mplus “complex design” option to control the nesting of students within schools and adjust standard errors for this clustering. When only the 5 ASC items were included (M1), the fit of the one-factor model fit was very good (e.g., CFI=0.977, TLI=0.955; Table 1). In addition, the ASC factor was highly reliable ($\alpha=0.87$, Omega=0.87) and well-defined (standardized factor loadings 0.72–0.83). Model M2 included all the study variables and provided correlations among the variables (see Results section). However, we note that the fit of this expanded model was also very good (e.g., CFI=0.977, TLI=0.949; Table 1).

Compositional Effects Models (M3-M6)

Our primary focus was on the compositional effects of L2Ach and L2SES on subsequent outcomes (Models M3-M6). In Models 3–5, ASC, GPA, and the three age-26 variables were considered as outcomes (see Fig. 1A). The interrelations among the five outcome variables were modeled as correlations, not in terms of effects of ASC and GPA on the three long-term outcomes. In this way, we estimated the effects of achievement and SES on the three long-term outcomes without controlling ASC and GPA. Thus, these models evaluate effects of L1Ach and L2Ach (M3), effects of L1SES and L2SES (M4), and the combined effects of all four variables (L1Ach and L2Ach, L1SES, and L2 SES; M5). For each set of models, we evaluated the effects with and without student-level demographic control variables (see Supplemental Materials, Sect. 6 for further discussion).

In the final model (M6), we repeated the analyses of compositional effects on the long-term outcomes while considering ASC and GPA as mediators (see Fig. 1B). In this way, we controlled the effects of ASC and GPA on the long-term outcomes. More specifically, we evaluated the total, direct, and indirect (mediated) effects of these models' L1 and L2 achievement and SES on the three long-term outcomes. Although the ten compositional effect models differ substantially in terms of degrees of freedom, the goodness-of-fit statistics are consistently excellent and highly similar across all the models (e.g., CFIs vary from 0.974 to 0.975 for Models M3-M6; Table 1).

Results

Correlations Among Individual-Level Student Variables

We present the correlations among all seven student-level (L1) variables (L1Ach, L1SES, ASC, GPA, and the three long-term outcomes). We based these correlations on the confirmatory-factor-analysis measurement model (M2 in Table 2). The seven variables are all positively correlated ($r_s = 0.20$ to 0.62). However, L1Ach, compared to L1SES, is more highly correlated with ASC (0.39 vs. 0.22) and particularly GPA (0.62 vs. 0.36). The three long-term outcomes correlated substantially with L1Ach (0.30 to 0.51) and GPA (0.29 to 0.56). These are higher than the corresponding correlations with L1SES (0.21 to 0.38) and ASC (0.20 to 0.32).

It is also relevant to note that both school-average variables (L2Ach and L2 SES) correlate substantially with all student-level (L1) predictor and outcome variables. Not surprisingly, L2Ach correlates most highly with L1Ach, and L2SES correlates most highly with L1SES. Nevertheless, L2Ach and L2SES also correlate positively with GPA and the three long-term outcomes. Thus, students in selective schools (with high SES and high achievement) tend to have better outcomes when not controlling for other variables. The critical question is how the size and direction of these

Table 2 Correlations among study variables

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Main predictor variables															
1. Academic self-concept	1														
2. Achievement	.39	1													
3. Socioeconomic status	.22	.50	1												
Demographic control variables															
4. Gender (female)	.07	.00	-.02	1											
5. Age	-.11	-.23	-.13	-.10	1										
6. Academic track	.25	.31	.21	.05	-.10	1									
7. Ethnicity-Black	.00	-.27	-.13	.01	.05	-.02	1								
8. Ethnicity-Hispanic	-.04	-.23	-.23	.00	.04	-.06	-.17	1							
9. Risk	-.17	-.39	-.38	-.02	.24	-.15	.23	.17	1						
End of high school															
10. Grade point average	.36	.62	.34	.18	-.21	.25	-.22	-.17	-.36	1					
Long-term (age 26) outcomes															
11. Occupational expectation	.20	.30	.21	.11	-.13	.17	-.03	-.04	-.17	.29	1				
12. Educational attainment	.31	.51	.38	.12	-.18	.25	-.11	-.14	-.31	.56	.38	1			
13. Educational expectation	.32	.48	.34	.10	-.18	.25	-.03	-.09	-.24	.44	.45	.64	1		
School-average predictors															
14. Achievement	.17	.58	.45	.01	-.11	.23	-.24	-.22	-.29	.31	.17	.32	.27	1.00	
15. Socioeconomic status	.17	.46	.57	.00	-.10	.24	-.15	-.22	-.27	.26	.19	.33	.29	.78	1.00

Correlations are based on Model M2 (Table 2). School-average predictors are based on aggregates of individual-student level variables. Because of the large sample size ($N = 16,197$), all $|r| > .01$ are statistically significant.

relations will change in the compositional models that control individual-student achievement and SES, as well as demographic variables.

The role of demographic variables in our study is primarily to control for pre-existing differences. Nevertheless, the size and direction of these relations are substantively interesting. Correlations among the six demographic control variables were mainly small, although most were statistically significant due to the substantial sample size. Gender differences also tended to be small. However, compared to boys, girls had higher ASCs, GPAs, and long-term outcomes (but did not differ on L1Ach). They were also more likely to be in an academic track and tended to be younger. Black and Hispanic students tended to have lower L1Ach, L1SES, GPAs, and educational attainment, but higher risk scores. Younger students had higher values on most outcomes (ASC, L1Ach, L1SES, GPA, and age-26 outcomes). However, ethnicity differences were small for ASC, occupational expectations, and educational expectations.

The largest correlations among the demographic variables involved the risk factor (age, $r=0.24$; academic track, -0.15 ; ethnicity-Black, 0.23 ; and ethnicity-Hispanic, 0.17). The correlation with age follows from the definition of the risk composite because repeating a year in school was one of the risk factors included in the composite. However, the risk factor was even more highly correlated with L1Ach (-0.39), L1SES (-0.38), and GPA (-0.36) and also correlated with the three long-term outcomes. Thus, it is important to include the composite risk factor in controlling for preexisting differences.

Compositional Effects of School-Average Achievement and SES

In this section, we specifically emphasize the results from the most comprehensive model M5, which includes student achievement (L1Ach and L2Ach), SES (L1SES and L2SES), and the six demographic control variables (Table 3; also see Fig. 1A). Our main focus is on school-compositional effects (Hypotheses 1 and 2). However, the models of these effects also consider the corresponding L1 effects (Lüdtke et al., 2008, 2011; see Supplemental Materials, Sect. 3 for a more detailed presentation of L1 effects). It is also relevant to compare Model M5 with the models that estimated the effects of each school-average variable separately (L2Ach in M3, L2SES in M4; Table 3) and with the models not controlling for demographic variables (Supplemental Materials Sect. 6).

Hypothesis 1: Effects of School-Average Achievement

We evaluated the compositional effects of L2Ach in two models; one that did not include L2SES (M3) and one that did (M5; see Table 3). In both models, L1Ach had consistently positive effects on all five outcomes. The effects of L1Ach were slightly larger in the models that did not include the demographic control variables (see Supplemental Materials, Sect. 6), indicating that it is important to control for these variables.

Table 3 Effects of individual-student level (L1) predictors and standardized compositional effects of school-average achievement (L2-Ach) and school-average SES (L2-SES)

Outcome	Predictors	M3: L2Ach (no L2SES)		M4: L2SES (no L2Ach)		M5: L2Ach & L2SES)	
		B	SE	B	SE	B	SE
ASC	L1Ach	.38	.01	.36	.01	.38	.01
	L1SES	.02	.01	.02	.01	.02	.01
GPA	L2Ach	-.19	.05			-.23	.07
	L2SES			.05	.04	.11	.05
	L1Ach	.48	.01	.47	.01	.48	.01
	L1SES	.03	.01	.03	.01	.04	.01
	ASC	.12	.01	.12	.01	.12	.01
Occ Exp	L2Ach	-.32	.06			-.23	.08
	L2SES			-.13	.04	-.07	.06
	L1Ach	.22	.01	.21	.01	.22	.01
Educ Exp	L1SES	.07	.01	.05	.01	.06	.01
	ASC	.06	.01	.06	.01	.06	.01
	L2Ach	-.08	.06			-.16	.07
Attain	L2SES			.14	.04	.13	.06
	L1Ach	.34	.01	.33	.01	.34	.01
	L1SES	.12	.01	.10	.01	.10	.01
Attain	ASC	.12	.01	.12	.01	.12	.01
	L2Ach	-.04	.05			-.16	.06
	L2SES			.21	.03	.21	.05
Attain	L1Ach	.32	.01	.32	.01	.32	.01
	L1SES	.13	.01	.11	.01	.11	.01
	ASC	.1	.01	.10	.01	.10	.01
Attain	L2Ach	.05	.05			-.04	.06
	L2SES			.2	.03	.14	.05

Values in bold are statistically significant ($p < .05$)

L1 individual-student level, *L2* school level, *ASC* Academic self-concept, *GPA* grade-point average at the end of high school, *SES* socioeconomic status (L1SES & L2SES), *Ach* achievement (standardized test scores, L1Ach & L2Ach), *Occ Exp* & *Educ Exp* Occupational and educational expectations (completed at age 26), *Attain* educational attainment (at age 26)

Compositional effects (shaded in grey) are the standardized effects of compositional variables (L2SES and L2Ach) on subsequent outcomes (see Fig. 1) based on multilevel doubly-latent models. Models differ in terms of the inclusion of L2ACH (Model M3), L2SES (Model M4), and both L2Ach and L2SES (Model M5; see Table 2 and Fig. 1). All models also control the effect of demographic variables (see Table 2; see Supplemental Materials Tables 2 and 3 for models without demographic control variables)

The comprehensive Model M5 included both L2ACH and L2SES. In this model, L2Ach negatively predicted 4 of the 5 outcomes (Table 3, shaded in grey). The largest effects were for ASC (-0.23) and GPA (-0.23), followed by age-26 educational expectations (-0.16) and occupational expectations (-0.16). However, the effect on age-26 educational attainment was not statistically significant. These results provide partial support for Hypothesis 1.

Hypothesis 2: Effects of School-Average SES

We evaluated the compositional effects of L2SES in two models; one that did not include L2Ach (M4) and one that did (M5; see Table 3). The effects of L1SES were consistently small but significantly positive for all five outcomes. However, the effects of L1SES were substantially larger in the models without demographic control variables (see Supplemental Materials), reflecting the substantial overlap between L1SES and the outcome variables. This finding confirms the importance of controlling for L1SES.

In the comprehensive Model M5, L2SES positively predicted 4 of 5 outcomes (shaded in grey in Table 3): ASC (0.11) and the three age-26 outcomes (educational attainment, 0.14; occupational expectations, 0.13; and educational expectations, 0.21). The effect on GPA was not significant. As such, the results provide partial support for Hypothesis 2.

Hypothesis 3: Mediation of Effects on Long-Term Outcomes

In Hypothesis 3, we posited that the effects of achievement and SES (L1Ach, L2Ach, L1SES, L2SES) on the long-term outcomes are mediated in part by ASC and GPA. The final model (M6) tested this hypothesis. In this model, we considered ASC and GPA as mediators of the effects of achievement and SES on the three long-term outcomes. We estimated the effects of ASC on GPA and the effects of both ASC and GPA on the long-term outcomes, as depicted in Fig. 1B. Model M6's rationale was to evaluate the extent to which the effects of L1 and L2 achievement and SES on long-term outcomes change when controlling for ASC and GPA. For each of the effects of L1 and L2 achievement and SES on long-term outcomes, we evaluated total effects, mediated effects (via ASC and GPA), and direct (unmediated) effects.

Student-Level (L1) Effects In Model M6, both L1Ach and L1SES have significantly positive total, direct, and indirect effects on all three long-term outcomes. The indirect effects of L1Ach are mediated through GPA and through ASC via GPA (i.e., ASC effects on long-term outcomes mediated by GPA). These results support Hypothesis 3. We also considered the total effects, which are the sum of all direct and indirect effects. For all three long-term outcomes, the total and direct effects of L1ACH are systematically larger than those for L1SES. For effects of L1Ach on the three outcomes, all total effects (0.24, 0.39, and 0.36), direct effects (0.15, 0.24, and 0.13), and mediated effects (0.09, 0.15, and 0.22) are statistically significant and substantial. The indirect effects are primarily mediated by GPA (0.07, 0.11, and 0.18). However, they are also mediated through ASC and through ASC via GPA. For L1SES, the total effects (0.06, 0.10, and 0.12), direct effects (0.05, 0.09, and 0.10), and mediated effects (0.01, 0.01, and 0.02) are all statistically significant, but smaller than those for L1Ach. Furthermore, unlike L1Ach, most of the effects of L1SES are direct effects.

Compositional Effects The total school-average compositional effects of L2Ach are negative for all three long-term outcomes (but nonsignificant for educational attainment; Table 4). In contrast, the compositional effects of L2SES are positive for all three outcomes (although nonsignificant for attainment).

For L2Ach, indirect effects were primarily mediated through GPA. These mediated effects were significantly negative for all three long-term outcomes (-0.06 , -0.10 , and -0.15 ; Table 4). However, indirect effects of L2Ach were also mediated through ASC via GPA. Although statistically significant and negative, the effects mediated through ASC and GPA were smaller in size than those mediated through GPA alone. The results supported Hypothesis 3, but the pattern of mediation varied across the three age-26 outcomes. For occupational and education expectations, total effects, direct effects, and total indirect effects were all negative. These indirect effects were mediated primarily through GPA. In contrast, for attainment, the total effects were nonsignificant. These were driven by a significant positive direct effect and a larger negative indirect effect mediated primarily by GPA (but also by ASC).

For L2 SES, indirect compositional effects on the three outcomes mediated through GPA were very small (-0.00 , -0.01 , and -0.02) but significant for educational expectations and attainment. The indirect effects mediated by ASC via GPA were also very small and only significant for educational expectations (-0.01). Most of the effects of L2SES were direct, unmediated effects.

Discussion

Our overarching purpose was to juxtapose the school-compositional effects of L2Ach and L2SES on ASC, GPA, and long-term outcomes at age 26. At the individual-student level, L1Ach, L1SES, and ASC were all significantly correlated to each other and subsequent outcomes (GPA and the three long-term outcomes). However, at the school-average level, the total effects of L2Ach were consistently adverse, whereas the total effects of L2SES were consistently positive. These results support our *a priori* predictions. They are also consistent with and extend Göllner et al.'s (2018) highly controversial conclusion that the optimal combination to maximize benefits for a student is a school with high L2SES but modest L2Ach.

Our results have important implications for understanding school selectivity based on L2Ach and L2SES. Parents, policymakers, and some researchers assume that placing a child in a highly selective school will improve the child's future success—in addition to the many preexisting advantages of students typically attending selective schools. However, this conventional wisdom is difficult to test because the preexisting differences inevitably bias results in favor of selective schools. Moreover, these differences can generate phantom effects that are difficult (or impossible) to control fully in observation studies.

Table 4 Multilevel mediated effects of achievement and SES on long-term outcomes: total, indirect (mediated), and direct effects (see Fig. 1B)

Effects	Level of analysis			
	Student (L1)		School-average (L2)	
	Est	SE	Est	SE
Effects: ACH to Occ Exp				
Total ^a	.24	.013	−.13	.064
Total indirect	.09	.008	−.05	.032
via ASC	.01	.005	.04	.031
via GPA	.07	.007	−.08	.014
via ASC via GPA	.01	.001	−.01	.003
Direct	.15	.016	−.09	.071
Effects: SES to Occ Exp				
Total ^a	.06	.012	.18	.056
Total indirect	.01	.001	.04	.023
via ASC	.00	.001	.04	.023
via GPA	.01	.001	.00	.003
via ASC via GPA	.00	.000	.00	.002
Direct	.05	.012	.14	.059
Effects: ACH to Educ Exp				
Total ^a	.39	.012	−.17	.056
Total indirect	.15	.008	−.11	.029
via ASC	.03	.005	.00	.026
via GPA	.11	.006	−.10	.013
via ASC via GPA	.01	.001	−.01	.002
Direct	.24	.015	−.06	.063
Effects: SES to Educ Exp				
Total ^a	.10	.010	.24	.047
Total indirect	.01	.002	.02	.020
via ASC	.00	.001	.03	.020
via GPA	.01	.002	−.01	.003
via ASC via GPA	.00	.000	.00	.002
Direct	.09	.010	.22	.048
Effects: ACH to attainment				
Total ^a	.36	.011	−.05	.056
Total Indirect	.22	.007	−.17	.026
via ASC	.02	.004	.00	.022
via GPA	.18	.006	−.15	.016
via ASC via GPA	.02	.001	−.01	.005
Direct	.13	.013	.11	.059
Effects: SES to attainment				
Total ^a	.12	.010	.16	.046
Total indirect	.02	.003	.00	.021
via ASC	.00	.001	.02	.019
via GPA	.01	.003	−.02	.009

Table 4 (continued)

Effects	Level of analysis			
	Student (L1)		School-average (L2)	
	Est	SE	Est	SE
via ASC via GPA	.00	.001	.00	.004
Direct	.10	.009	.16	.048

Values in bold are statistically significant ($p < .05$)

ASC Academic self-concept, SES socioeconomic status, ACH achievement (standardized test scores), GPA final grade point average at the end of high school, Occ Exp & Educ Exp occupational and educational expectations (completed at age 26), Attain educational attainment (at age 26).

^aIn supplemental analyses, we showed that differences between the negative compositional effect of school-average achievement and the positive compositional effect of school-average SES were significant for each of the three long-term outcomes (all $p < .05$).

Support for Alternative Interpretations of Recent Compositional Studies

Our study supports a growing consensus concerning appropriate methodology, theory, and empirical conclusions. Methodologically, we used the doubly-latent multilevel compositional SEM with appropriate control for covariates. This is widely acknowledged as best practice (e.g., Becker et al., 2022; Dicke et al., 2018; Göllner et al., 2018; Lüdtke et al., 2008, 2011; Televantou et al., 2015, 2021). Theoretically, our study supports the classic distinction between assimilation and contrast effects (Kelley, 1952; Suls & Wheeler, 2000). This distinction leads to predictions that students identify with other students in high L2SES schools (assimilation or reflected glory effect) but contrast themselves with other students in high L2Ach schools (Marsh & O'Mara, 2010). Empirically, our study adds to the growing number of studies supporting the robustness of the negative effect of L2Ach on ASC, the BFLPE. However, other aspects of our research are more controversial, including issues like phantom effects that are particularly relevant to recent school compositional studies (e.g., Becker et al., 2022; Göllner et al., 2018; von Keyserlingk et al., 2020).

Phantom Effects: Failure to Control Preexisting Differences

The adverse effects of L2Ach on ASC have a robust theoretical and empirical basis. However, the corresponding effects of L2Ach on other achievement-related outcomes are highly contested. Indeed, as noted earlier, Harker and Tymms (2004) referred to the so-called positive effects of L2Ach on subsequent L1Ach as “phantom effects” that disappear with appropriate control for measurement error and

covariates. This interpretation is consistent with several recent compositional studies based on doubly-latent multilevel SEMs. These show that L2Ach effects on subsequent achievement tend to be zero or even negative when appropriate controls are included (Dicke et al., 2018; Televantou et al., 2015, 2021). However, Becker et al. (2022) argued that their results countered the claim that the positive effects of L2Ach were merely phantom effects due to methodological issues.

The Becker et al. (2022) study challenges our conclusion about L2Ach's negative effects. However, the role of track is a critical issue in Becker et al.'s research based on German secondary schools. In these schools, explicit tracking at the school level is determined mainly by school performance in primary school before students begin high school (see Marsh et al., 2018). Thus, track reflects a cumulative measure of performance in primary school. Furthermore, it is influenced by cognitive and noncognitive variables distinct from standardized achievement tests (e.g., motivation and conscientious; see discussion by Borghans et al., 2016). Hence, track controls preexisting differences beyond those associated with test scores used in most studies. However, high-track schools in the German system also reflect better resourcing and an advanced curriculum. As such, track is a crude (dichotomous) measure of some combination of prior achievement, noncognitive variables, and current resourcing.

Becker et al. found that controlling track substantially reduced the positive effects of L2Ach. In one case, these effects even became significantly negative. Concerning peer spillover effects that were a major focus of Becker et al.'s and our study, both preexisting differences in achievement and current resourcing differences can generate positive biases. Hence, the Becker et al. results are consistent with a phantom-effect interpretation of peer spill-over effects. The apparently positive effects of L2Ach are substantially reduced and might disappear altogether—or even become negative—with better controls. However, this role of track is somewhat idiosyncratic to the German system, where track is a rigidly defined category of school type rather than a loosely defined measure of within-school tracking as in the ELS:2002 database. Nevertheless, there is a need for further research that more effectively distinguishes the effects of pre-existing differences in achievement and noncognitive variables, resourcing, and curriculum on peer spill-over effects.

More broadly, it seems likely that all estimated school-composition effects are confounded substantially by preexisting differences. These are inevitably under-controlled, thus at least in part generating phantom effects. Moreover, because the biases generated by preexisting differences are so strong, it is unlikely that phantom effects can ever be eliminated entirely in observational studies, no matter what covariates are available. Hence, it is a matter of how large these biases are relative to observed effects and to the strength of controls for preexisting differences.

However, the implications of these inevitable biases differ fundamentally for negative contrast effects and positive assimilation effects. For contrast effects, L2Ach effects are predicted to be negative (like the negative effects of L2Ach reported here). The positive bias due to preexisting differences is conservative concerning this prediction (i.e., the bias works opposite to the prediction). On the other hand, for positive assimilation effects, preexisting differences positively bias the results (i.e., the

bias is in the same direction as the prediction). Because predicted assimilation effects are confounded with preexisting differences, it is inevitable that some (or, perhaps, even all) observed assimilation effects are due to this bias (i.e., they are due at least in part to phantom effects). Consistent with this perspective, both Dicke et al. (2018) and Televantou et al. (2021) showed that BFLPEs for ASCs were conservative in relation to these biases; they became more negative with control for measurement error and covariates, including prior achievement. Conversely, apparently positive effects of L2Ach on subsequent individual achievement disappeared or became significantly negative with control for measurement error and covariates.

Implicit in the Becker et al. (2022) interpretation of positive L2Ach effects is the suggestion that there are benefits associated with ability stratification and explicit tracking—at least for students in high-achieving schools. However, it is crucial to consider this issue in the broader research context on the relation between academic excellence and inequality. Based on five cycles of PISA assessments (PISA2000 to PISA2012) for 27 OECD countries, Parker et al. (2018) showed that countries with greater ability stratification had lower average student achievement. Furthermore, Parker et al. also evaluated the effects of changes in ability stratification over time for each country. These results showed that countries with increasing ability stratification had decreasing levels of achievement. The adverse effects of ability stratification were particularly evident for low- and average-achieving students. Thus, county-level inequality associated with ability stratification is negatively related to excellence based on achievement.

Juxtaposing School-Average Achievement and School-Average SES

Göllner et al. (2018) emphasized how L2SES contributes positively to students' long-term success. Their rationale is similar to the arguments by Becker et al. (2022) and many others regarding school selectivity based on achievement. Indeed, Göllner et al. distinguished between “good schools” in terms of school facilities, including better teachers and resources (like Becker et al.'s “instructional processes”) and contagion (like Becker et al.'s positive peer spillover effects). However, Göllner et al. suggested that school composition studies rarely consider L2SES and L2Ach in the same model (but see earlier discussion of Marsh, 1991; Marsh & O'Mara, 2010). As described earlier, Göllner et al. used historical archive data from the 1960s to show that L2SES effects were positive but corresponding L2Ach effects were adverse. Göllner et al.'s highly controversial conclusion was that the optimal combination is schools with high L2SES but lower L2Ach. Nevertheless, they noted caution given their use of historical data and inherent difficulties in disentangling the effects of L2SES and L2Ach and called for further research.

Our results are consistent with Göllner et al.'s (2018) highly provocative interpretation, juxtaposing the benefits of L2SES and the adverse effects of L2Ach. A unique aspect of Göllner et al.'s results is access to 50-year follow-up data. Nevertheless, our results are stronger in many ways (more recent, less attrition, better controls for missing data, more robust demographic control variables, and the inclusion

of ASC and GPA as mediating variables). In this respect, the studies complement each other, demonstrating the need to consider both L2Ach and L2SES in school-composition studies.

Strengths, Weaknesses, and Directions for Further Research

Particular strengths of our study are the large, nationally representative ELS:2002 database, the inclusion of final high school GPA based on official school transcripts, and the age-26 outcomes collected following the postschool transition into early adulthood. Methodologically, we applied doubly-latent multilevel SEMs with 20 multiple imputation data sets based on extensive auxiliary variables to control missing data and strong covariates to control preexisting differences. Although routinely used in BFLPE research and increasingly used in L2Ach composition studies, doubly-latent modeling is rare in studies juxtaposing the effects of L2SES and L2Ach (as also emphasized by Göllner et al., 2018). Furthermore, many school compositional studies were cross-sectional, and few included long-term outcomes as well as multiple waves of high school outcomes. Our study is a substantive-methodological synergy and has important policy, practice, and parental choice implications.

There are also potentially important limitations to our study. As with all correlational studies, support for *a priori* hypotheses that imply causality must be interpreted cautiously. However, the most important threat to causal interpretations is the lack of control for potential covariates that are confounded with compositional effects. Here, we considered a robust set of demographic control variables (gender, age, SES, achievement, track, and the composite risk variable). However, their inclusion had relatively little impact on the pattern of results in support of our *a priori* predictions (see Supplemental Materials, Sect. 6). Nevertheless, a direction for further research is testing the extent to which school-compositional effects generalize over subgroups based on demographic control variables.

ELS:2002's initial wave of data is almost 20 years old, and even the final wave of long-term outcomes was collected ten years ago. Although somewhat dated, the findings contribute to a well-established historical pattern of results, based primarily on US data, that show positive L2SES effects but adverse L2Ach effects. These results were evident in large, nationally representative samples in the early 1960s (Project "TALENT", Göllner, et al., 2018; also see reviews by Alwin & Otto, 1977), late 1960s (Youth in Transition study, Marsh & O'Mara, 2010; also see Bachman & O'Malley, 1986; Marsh, 1987), 1980s (High School and Beyond study, Marsh, 1991), and 2000s (the current study).

Generalizability of School Composition Effects

We agree with Becker et al. (2022) that the divergence of findings concerning school-composition effects is due only partly to methodological issues. Becker et al. argued that systematic reviews and further research are needed to evaluate the settings and

conditions that lead to different effects and “ultimately, which settings may be conducive to offering maximum student benefit” (p. 14). Progress requires substantive-methodological synergy. Disentangling these competing interpretations requires more detailed data and theory about mediating mechanisms. Future research needs to include actual pretest data from before the start of high school to control the inherent bias in favor of selective schools. It will also be essential to include variables specifically designed to differentiate the posited effects of assimilation (reflected glory and positive peer spillover effects; e.g., Marsh et al., 2000; Trautwein et al., 2006), contrast (social comparison effects; e.g., Huguët, et al., 2009; Marsh et al., 2014), and resources (e.g., expenditure, school facilities, curriculum, class size, and teacher qualifications; Becker et al., 2022; Hattie, 2002).

We also note that the size and direction of school compositional effects will vary substantially across different outcomes. For example, the effects of L2Ach are more negative for ASC, but less negative, nonsignificant, or even positive for achievement. Even for studies more narrowly focused on test scores as outcomes, the match between the curriculum and the tests is likely to be critical. Thus, for example, if high-track students study more advanced material and this material is the basis of the tests, L2Ach effects are likely to be more positive than for tests based on materials common to all the tracks.

Generalizability of School Composition Effects on Mental Health and Nonacademic Outcomes

The research program by Luthar and colleagues demonstrates the adverse effects of attending high-achieving schools on student mental health (anxiety, depression, distress, delinquency, substance abuse, high-risk behaviors, and adverse childhood experiences e.g., Ebbert et al., 2019; Luthar & Kumar, 2018; Luthar et al., 2020). Luthar (2003), Luthar and Ansary (2005), and Luthar and Latendresse (2005) initially identified seemingly paradoxical increased risks of psychological problems for students from affluent families (“affluenza”). However, subsequent large-scale multilevel studies by Coley et al. (2018; also see Lund & Dearing, 2012; Lund et al., 2017) showed that these effects on mental health problems were due to school compositional effects rather than effects of L1 family SES. This led Luthar and colleagues to shift from individual-student characteristics to an emphasis on high-achieving schools (e.g., Ebbert et al., 2019; Luthar & Kuman, 2018; Luthar et al., 2020). They also emphasized the importance of a robust self-concept to children’s mental health, which can be compromised in high-achieving schools where self-worth is based on relative accomplishments and social comparison.

Luthar et al.’s (2020) research program complements the research presented here in many ways. Both highlight seemingly paradoxically negative effects of attending high-achieving schools, driven by social comparison processes. In addition, both

emphasize important public policy implications for parents, schools, and social policy, developmental perspectives, and multilevel ecological approaches. Interestingly, however, there is surprisingly little cross-citation of the academic outcome studies reviewed here and the mental health research by Luthar and colleagues. The major exception is Luthar et al.'s (2020) discussion of the happy-fish-little-pond effect (Pekrun et al., 2019), based on the application of the BFLPE to emotions (rather than ASC). In addition, Luthar et al. (2020) cited Göllner et al. (2018) as showing that affluent high-achieving schools were associated with poorer long-term educational and occupational outcomes. This is critical as Göllner et al.'s study is an essential basis of our research.

However, the Luthar et al. (2020) conceptual model goes beyond contrast effects driven by social comparison processes posited in the BFLPE model. They emphasize the pressures to achieve in high-achieving schools (e.g., expectations of parents and teachers, student envy, perfectionistic tendencies, and competition to gain acceptance to top universities) and potential interventions to counteract the negative effects of high-achieving schools. Their research, like ours, also demonstrates the importance of unconfounding school-level effects from the effects of individual-student characteristics. Nevertheless, their research does not fully resolve whether the negative effects of high-achieving schools are driven by L2Ach (which remains implicit in their model) or by L2SES. Indeed, L2SES rather than L2Ach was the basis of the Coley et al. (2018; Lund & Dearing, 2012; Lund et al., 2017) studies which had prompted Luthar et al. to shift from a focus on L1SES to focusing on high-achieving schools. As shown in the present investigation, the distinction between L2Ach and L2SES is critical and has important substantive and theoretical implications. More broadly, it will be important for future research to more fully integrate the strengths of these complementary research programs.

Cross-National Generalizability

Keyserlingk et al. (2020) recognized the need for cross-national comparisons to test the generalizability of school composition effects. We agree that there is a need for cross-national studies to evaluate better the generalizability of school composition effects and the conditions under which they vary. More broadly, cross-national generalizability is an important macrolevel issue. A major limitation of much educational research is overreliance on studies from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies (Hendriks et al., 2019)—particularly the US and a few other industrialized countries. This limitation also undermines the generalizability of results based on systematic reviews and meta-analyses based mainly on studies from WEIRD countries (see discussion by Marsh et al., 2020). Although evident in most areas of educational research, this issue is particularly relevant for the studies considered here, given that these studies are based primarily on US and German samples. We illustrated a cross-national approach, demonstrating the cross-national generalizability of the BFLPE

based on data from the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study. Nevertheless, these databases' cross-national (single wave) nature is a major limitation in disentangling school-composition effects from the effects of preexisting differences—particularly prior achievement.

Conclusions and Implications

Our intent is to change conventional wisdom about the effects of L2Ach and how educational psychologists study these constructs. Our substantive-methodological synergy brings together strong data, methodological models, and theory to address substantive issues with important consequences for policy and practice—a substantive-methodological synergy. The issues at the heart of our research have critical implications for parents and policy. For example, parents must choose the schools their children attend and even uproot their families to live in areas with “good” schools. In addition, policymakers seek to allocate students to schools to maximize benefits for all students. For example, good schools are often characterized by those with high levels of L2SES or L2Ach. However, there is limited research juxtaposing the effects of the two compositional effects. In addressing this issue, we replicate and extend Göllner et al.'s (2018) highly controversial conclusion that the optimal balance for a good school is a high level of L2SES but a moderate or low level of L2Ach.

There is universal support for the finding that L2Ach has adverse effects on ASC (the BFLPE) and related psychosocial variables (e.g., aspirations, interests, and emotions). Here, we extend this research. We replicate Göllner et al.'s (2018) finding that L2Ach also negatively affects long-term outcomes in later life. Also, we found that the negative effects of L2Ach became more negative after controlling for SES at the individual-student and school-average levels. This suggests that researchers need to consider both compositional effects simultaneously to understand each better. However, there is also a need for stronger theoretical models to explain why the effects of L2Ach become more negative after controlling for L2SES. In contrast, the positive effects of L2SES are less affected by controlling L2Ach.

In summary, our results and research review suggest negative effects associated with L2Ach but positive effects related to L2SES. However, current research—including our study—has not adequately disentangled these compositional effects from competing effects. These include resourcing effects (spending, school facilities, curriculum, class size, teacher qualifications, etc.), assimilation effects (reflected glory and positive peer spillover effects), and contrast effects (social comparison processes, BFLPEs, and negative peer spillover effects). From this perspective, we agree with Becker et al. (2022) that we need to stop looking for universal conclusions about school compositional effects. Instead, future research needs to focus on stronger theoretical models underpinning school-composition effects and the conditions and circumstances that maximize student benefits.

Appendix

Table 5 Glossary of abbreviations used in the present investigation

Abbreviation	Definition of the abbreviation
ASC	Academic self-concept
Ach	Achievement based on test scores
L1Ach	Individual student achievement
L2Ach2	School-average achievement (aggregate of L1 achievement)
BFLPE	Big-fish-little-pond effect
CFI	Comparative fit index of goodness of fit
ELS:2002	Educational Longitudinal Survey of 2002
GPA	Grade point average
SEMs	Structural equation models
SES	Socioeconomic status
L1SES	Socioeconomic status at the level of the individual student
L2SES	School-average socioeconomic status
TLI	Tucker-Lewis index of goodness-of-fit
Track	Academic track (1 = academic, 0 = nonacademic)
WEIRD	Western, Educated, Industrialized, Rich, and Democratic

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10648-023-09726-4>.

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

Data Availability See ELS:2002 website for study design, variables, access to data used here and studies using these data (<https://nces.ed.gov/surveys/els2002/>)

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alicke, M. D., Zell, E., & Bloom, D. L. (2010). Mere categorization and the frog-pond effect. *Psychological Science*, 21, 174–177. <https://doi.org/10.1177/0956797609357718>
- Alwin, D. F., & Otto, L. B. (1977). High school context effects on aspirations. *Sociology of Education*, 50, 259–273. <https://doi.org/10.2307/2112499>

- Asparouhov, T., & Muthén, B. M. (2010). Resampling methods in Mplus for complex survey data. http://statmodel.com/download/Resampling_Methods5.pdf. Accessed 22 May 2023
- Bachman, J. G., & O'Malley, P. M. (1986). Self-concepts, self-esteem, and educational experiences: The frogpond revisited (again). *Journal of Personality and Social Psychology*, 50, 33–46. <https://doi.org/10.1037/0022-3514.50.1.35>
- Becker, M., Kocaj, A., Jansen, M., Dumont, H., & Lüdtke, O. (2022). Class-average achievement and individual achievement development: Testing achievement composition and peer spillover effects using five German longitudinal studies. *Journal of Educational Psychology*, 114(1), 177–197. <https://doi.org/10.1037/edu0000519>
- Borghans, L., Golsteyn, B. H., Heckman, J. J., & Humphries, J. E. (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences*, 113(47), 13354–13359.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Government Printing Office.
- Coleman, J. S., & Hoffer, T. (1987). *Public and private high schools: The impact of communities*. Basic Books.
- Coley, R. L., Sims, J., Dearing, E., & Spielvogel, B. (2018). Locating economic risks for adolescent mental and behavioral health: Poverty and affluence in families, neighborhoods, and schools. *Child Development*, 89(2), 360–369. <https://doi.org/10.1111/cdev.12771>
- Davis, J. A. (1966). The campus as a frog pond: An application of theory of relative deprivation to career decisions for college men. *American Journal of Sociology*, 72, 17–31. <https://doi.org/10.1086/224257>
- Dicke, T., Marsh, H. W., Parker, P. D., Pekrun, R., Guo, J., & Televantou, I. (2018). Effects of school-average achievement on individual self-concept and achievement: Unmasking phantom effects masquerading as true compositional effects. *Journal of Educational Psychology*, 110(8), 1112–1126. <https://doi.org/10.1037/edu0000259>
- Diener, E., & Fujita, F. (1997). Social comparison and subjective well-being. In B. P. Buunk & F. X. Gibbons (Eds.), *Health, coping, and well-being: Perspectives from social comparison theory* (pp. 329–358). Erlbaum.
- Ebbert, A. M., Kumar, N. L., & Luthar, S. S. (2019). Complexities in adjustment patterns among the “best and the brightest”: Risk and resilience in the context of high achieving schools. *Research in Human Development*, 16(1), 21–34. <https://doi.org/10.1080/15427609.2018.1541376>
- Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.
- Espenshade, T., Hale, L. E., & Chung, C. Y. (2005). The frog pond revisited. *Sociology of Education*, 78(4), 269–293.
- Fang, J., Huang, X., Zhang, M., Huang, F., Li, Z., & Yuan, Q. (2018). The big-fish-little-pond effect on academic self-concept: A meta-analysis. *Frontiers in Psychology*, 9, 1569. <https://doi.org/10.3389/fpsyg.2018.01569>
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117–140. <https://doi.org/10.1177/001872675400700202>
- Frank, R. H. (1985). *Choosing the right pond: Human behavior and the quest for status*. Oxford University Press.
- Frank, R. H. (2012). *The Darwin economy: Liberty, competition, and the common good*. Princeton University Press. <https://doi.org/10.1515/9781400844982>
- Göllner, R., Damian, R. I., Nagengast, B., Roberts, B. W., & Trautwein, U. (2018). It's not only who you are but who you are with: High school composition and individuals' attainment over the life course. *Psychological Science*, 29(11), 1785–1796. <https://doi.org/10.1177/0956797618794454>
- Graham, J. W. (2009). *Missing data analysis: Making it work in the real world*. *Annual Review of Psychology*, 60, 549–576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>
- Gutman, L. M., & Schoon, I. (2013). The impact of non-cognitive skills on outcomes for young people. A literature review. Institute of Education, University of London. https://discovery.ucl.ac.uk/id/eprint/10125763/1/Gutman_Schoon_%202013%20Non-cognitive_skills_literature_review_.pdf. Accessed 22 May 2023
- Greenwald, A. G. (1988). A social-cognitive account of the self's development. In D. K. Lapsley & F. C. Power (Eds.), *Self, ego and identity: Interpretative approaches* (pp. 30–42). New York: Springer-Verlag.
- Harker, R., & Tymms, P. (2004). The effects of student composition on school outcomes. *School Effectiveness and School Improvement*, 15(2), 177–199. <https://doi.org/10.1076/sesi.15.2.177.30432>

- Harris, D. N. (2010). How do school peers influence student educational outcomes? Theory and evidence from economics and other social sciences. *Teachers College Record*, 112, 1163–1197. <https://doi.org/10.1177/016146811011200404>
- Hattie, J. A. C. (2002). Classroom composition and peer effects. *International Journal of Educational Research*, 37(5), 449–481. [https://doi.org/10.1016/S0883-0355\(03\)00015-6](https://doi.org/10.1016/S0883-0355(03)00015-6).
- Hendriks, T., Warren, M. A., Schotanus-Dijkstra, M., Hassankhan, A., Graafsma, T., Bohlmeijer, E., & de Jong, J. (2019). How WEIRD are positive psychology interventions? A bibliometric analysis of randomized controlled trials on the science of well-being. *The Journal of Positive Psychology*, 14(4), 489–501. <https://doi.org/10.1080/17439760.2018.1484941>
- Heckman, J. J., & Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review*, 91(2), 145–149. <https://doi.org/10.1257/aer.91.2.145>
- Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 411–482.
- Helson, H. (1964). *Adaptation-level theory*. Harper & Row.
- Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, 49(5), 505–528. <https://doi.org/10.1016/j.jsp.2011.07.001>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>.
- Huguet, P., Dumas, F., Marsh, H., Régner, I., Wheeler, L., Suls, J., Seaton, M., & Nezlek, J. (2009). Clarifying the role of social comparison in the big-fish–little-pond effect (BFLPE): An integrative study. *Journal of Personality and Social Psychology*, 97(1), 156–170. <https://doi.org/10.1037/a0015558>
- Hyman, H. (1942). The psychology of subjective status. *Psychological Bulletin*, 39, 473–474.
- Ingels, S.J., Pratt, D.J., Alexander, C.P., Jewell, D.M., Lauff, E., Mattox, T.L., & Wilson, D. (2014). *Education Longitudinal Study of 2002: third follow-up data file documentation (NCES 2014–364)*. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Ingels, S.J., Pratt, D.J., Rogers, J.E., Siegel, P.H., & Stutts, E.S. (2004). *Education Longitudinal Study of 2002: base year data file user's manual (NCES 2004–405)*. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Ingels, S.J., Pratt, D.J., Rogers, J.E., Siegel, P.H., & Stutts, E.S. (2005). *Education Longitudinal Study of 2002: base-year to first follow-up data file documentation (NCES 2006–344)*. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Ingels, S.J., Pratt, D.J., Wilson, D., Burns, L.J., Currivan, D., Rogers, J.E., & HubbardBednasz, S. (2007). *Education Longitudinal Study of 2002: base-year to second follow-up data file documentation (NCES 2008–347)*. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- James, W. (1890/1963). *The principles of psychology* (Vol. 2). Holt. <https://doi.org/10.1037/10538-000>
- Jelčić, H., Phelps, E., & Lerner, R. M. (2009). Use of missing data methods in longitudinal studies: the persistence of bad practices in developmental psychology. *Developmental Psychology*, 45(4), 1195–1199. <https://doi.org/10.1037/a0015665>.
- Kelley, H. H. (1952). Two functions of reference groups. *Readings in Social Psychology*, 2, 410–414.
- Lüdtke, O., & Robitzsch, A. (2021). A critique of the random intercept cross-lagged panel model. *PsyArXiv*. <https://doi.org/10.31234/osf.io/6f85c>.
- Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, 13(3), 203–229. <https://doi.org/10.1037/a0012869>
- Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2x2 taxonomy of multilevel latent contextual models: Accuracy-bias trade-offs in full and partial error-correction models. *Psychological Methods*, 16(4), 444–467. <https://doi.org/10.1037/a0024376>
- Lund, T. J., & Dearing, E. (2012). Is growing up affluent risky for adolescents or is the problem growing up in an affluent neighborhood? *Journal of Research on Adolescence*. Advance online publication. <https://doi.org/10.1111/j.1532-7795.2012.00829.x>
- Lund, T. J., Dearing, E., & Zachrisson, H. D. (2017). Is affluence a risk for adolescents in Norway? *Journal of Research on Adolescence*, 27(3), 628–643.
- Luthar, S. S. (2003). The culture of affluence: Psychological costs of material wealth. *Child Development*, 74(6), 1581–1593. <https://doi.org/10.1046/j.1467-8624.2003.00625.x>

- Luthar, S. S., & Ansary, N. S. (2005). Dimensions of adolescent rebellion: Risks for academic failure among highand low-income youth. *Development and Psychopathology*, 17, 231–250. <https://doi.org/10.1017/S0954579405050121>
- Luthar, S. S., & Latendresse, S. J. (2005). Children of the affluent: Challenges to well-being. *Current Directions in Psychological Science*, 14(1), 49–53.
- Luthar, S. S., Kumar, N. L. (2018). Youth in high-achieving schools: Challenges to mental health and directions for evidence-based interventions. In: A. Leschied, D. Saklofske, G. Flett (Eds.), *Handbook of school-based mental health promotion. The springer series on human exceptionality*. Cham: Springer. https://doi.org/10.1007/978-3-319-89842-1_23
- Luthar, S. S., Kumar, N. L., & Zillmer, N. (2020). High-achieving schools connote risks for adolescents: Problems documented, processes implicated, and directions for interventions. *American Psychologist*, 75(7), 983–995. <https://doi.org/10.1037/amp0000556>
- Marsh, H. W. (1987). The big fish little pond effect on academic self-concept. *Journal of Educational Psychology*, 79(3), 280–295. <https://doi.org/10.1037/0022-0663.79.3.280>
- Marsh, H. W. (1991). The failure of high ability high schools to deliver academic benefits: The importance of ASC and educational aspirations. *American Educational Research Journal*, 28, 445–480. <https://doi.org/10.3102/00028312028002445>
- Marsh, H.W. (2007). *Self-concept theory, measurement and research into practice: the role of self concept in educational psychology – 25th Vernon-Wall lecture series*. British Psychological Society.
- Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, 1(2) 133–163. <https://doi.org/10.1111/j.1745-6916.2006.00010.x>.
- Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) Findings. *Structural Equation Modeling: A Multidisciplinary Journal*, 11(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2.
- Marsh, H. W., Köller, O., & Baumert, J. (2001). Reunification of East and West German school systems: Longitudinal multilevel modeling study of the big-fish-little-pond effect on academic self-concept. *American Educational Research Journal*, 38, 321–350.
- Marsh, H. W., Kong, C.-K., & Hau, K.-T. (2000). Longitudinal multilevel models of the big-fish-little-pond effect on academic self-concept: Counterbalancing contrast and reflected-glory effects in Hong Kong schools. *Journal of Personality and Social Psychology*, 78(2), 337–349. <https://doi.org/10.1037/0022-3514.78.2.337>
- Marsh, H. W., Kuyper, H., Morin, A. J. S., Parker, P. D., & Seaton, M. (2014). Big-fish-little-pond social comparison and local dominance effects: Integrating new statistical models, methodology, design, theory and substantive implications. *Learning and Instruction*, 33, 50–66. <https://doi.org/10.1016/j.learninstruc.2014.04.002>
- Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., & Köller, O. (2012a). Classroom climate and contextual effects: conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist*, 47(2), 106–124. <https://doi.org/10.1080/00461520.2012.670488>.
- Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, 44, 764–802. <https://doi.org/10.1080/00273170903333665>
- Marsh, H. W., & O'Mara, A. J. (2010). Long-term total negative effects of school-average ability on diverse educational outcomes. *Zeitschrift Für Pädagogische Psychologie/german Journal of Educational Psychology*, 24(1), 51–72. <https://doi.org/10.1024/1010-0652/a000004>
- Marsh, H. W., & Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, 47, 213–231. <https://doi.org/10.1037/0022-3514.47.1.213>
- Marsh, H. W., Parker, P. D., Guo, J., Basarkod, G., Niepel, C., & Van Zanden, B. (2021a). Illusory gender-equality paradox, math self-concept, and frame-of-reference effects: New integrative explanations for multiple paradoxes. *Journal of Personality and Social Psychology*, 121(1), 168–183. <https://doi.org/10.1037/pspp0000306>
- Marsh, H. W., Parker, P. D., Guo, J., Pekrun, R., & Basarkod, G. (2020). Psychological comparison processes and self-concept in relation to five distinct frame-of-reference effects: Pan-human

- cross-cultural generalizability over 68 countries. *European Journal of Personality*, 34, 180–202. <https://doi.org/10.1002/per.2232>
- Marsh, H. W., Pekrun, R., Murayama, K., Arens, A. K., Parker, P. D., Guo, J., & Dicke, T. (2018). An integrated model of academic self-concept development: Academic self-concept, grades, test scores, and tracking over 6 years. *Developmental Psychology*, 54, 263–280. <https://doi.org/10.1037/dev0000393>
- Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish-little-pond effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review*, 20, 319–350.
- Marsh, H. W., & Seaton, M. (2015). The big-fish-little-pond effect, competence self-perceptions, and relativity: substantive advances and methodological innovation. In: Elliott, A. J. (Ed.). *Advances in Motivation Science*. (vol. 2, 127–184). Elsevier. ISBN 9780128022702.
- Marsh, H. W., Xu, K. M., Parker, P. D., Hau, K.-T., Pekrun, R., Elliot, A., Guo, J., Dicke, T., & Basarkod, G. (2021b). Moderation of the big-fish-little-pond effect: Juxtaposition of evolutionary (Darwinian-economic) and achievement motivation theory predictions based on a Delphi approach. *Educational Psychology Review*, 33(4), 1353–1378. <https://doi.org/10.1007/s10648-020-09583-5>
- Marsh, H. W., Xu, M., & Martin, A. J. (2012b). Self-concept: a synergy of theory, method, and application. In: Harris, K. R., Graham, S., Urdan, T., McCormick, C. B., Sinatra, G. M., Sweller, J. (Eds.), *APA educational psychology handbook, Vol. 1. Theories, constructs, and critical issues* (pp. 427–458). <https://doi.org/10.1037/13273-015>
- Mayer, S. E., & Jencks, C. (1989). Growing up in poor neighborhoods: How much does it matter? *Science*, 243, 1441–1445. <https://doi.org/10.1126/science.243.4897.1441>
- Meyer, J. W. (1970). High school effects on college intentions. *American Journal of Sociology*, 76, 59–70. <https://doi.org/10.1086/224906>
- Morse, S., & Gergen, K. J. (1970). Social comparison, self-consistency, and the concept of self. *Journal of Personality & Social Psychology*, 16, 148–156. <https://doi.org/10.1037/h0029862>
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus: Statistical analysis with latent variables: User's Guide* (Version 8). Los Angeles, CA: Authors.
- Newman, D. A. (2014). Missing Data. *Organizational Research Methods*, 17(4), 372–411. <https://doi.org/10.1177/1094428114548590>
- Parducci, A. (1995). *Happiness, pleasure, and judgment: The contextual theory and its applications*. Erlbaum.
- Parker, P. D., Marsh, H. W., Jerrim, J., Guo, J., & Dicke, T. (2018). Trade-offs between equity and excellence in academic performance: Evidence from 27 OECD countries. *American Education Research Journal*, 55(4), 836–858. <https://doi.org/10.3102/0002831218760213>
- Parker, Dicke, T., Guo, J., Basarkod, G., & Marsh, H. (2021). Ability stratification predicts the size of the big-fish-little-pond effect. *Educational Researcher : A Publication of the American Educational Research Association*, 50(6), 334–344. <https://doi.org/10.3102/0013189X20986176>
- Pekrun, R., Murayama, K., Marsh, H. W., Goetz, T., & Frenzel, A. C. (2019). Happy fish in little ponds: Testing a reference group model of achievement and emotion. *Journal of Personality and Social Psychology*, 117(1), 166–185. <https://doi.org/10.1037/pspp0000230>
- Seaton, M., Marsh, H. W., & Craven, R. G. (2010). Big-fish-little-pond effect: Generalizability and moderation—two sides of the same coin. *American Educational Research Journal*, 47, 390–433. <https://doi.org/10.3102/0002831209350493>
- Sherif, M., & Sherif, C. W. (1969). *Social Psychology*. Harper & Row.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417–453. <https://doi.org/10.3102/00346543075003417>
- Stouffer, S.A., Suchman, E.A., DeVinney, L.C., Star, S.A. & Williams, R.M. (1949). *The American soldier: Adjustments during army life* (Vol. 1). Princeton University Press.
- Suls, J., Wheeler, L. (2000). A selective history of classic and neo-social comparison theory. In Suls, J., Wheeler, L. (eds) *Handbook of Social Comparison*. The Springer Series in Social Clinical Psychology. https://doi.org/10.1007/978-1-4615-4237-7_1
- Televantou, I., Marsh, H. W., Dicke, T., & Nicolaides, C. (2021). Phantom and big-fish-little-pond-effects on academic self-concept and academic achievement: Evidence from English early primary schools. *Learning and Instruction*, 71, 101399–101410. <https://doi.org/10.1016/j.learninstruc.2020.101399>
- Televantou, I., Marsh, H. W., Kyriakides, L., Nagengast, B., Fletcher, J., & Malmberg, L.-E. (2015). Phantom effects in school composition research: Consequences of failure to control biases due to measurement error in traditional multilevel models. *School Effectiveness and School Improvement*, 26(1), 75–101. <https://doi.org/10.1080/09243453.2013.871302>

- Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology*, 98(4), 788–806. <https://doi.org/10.1037/0022-0663.98.4.788>
- Upshaw, H. S. (1969). The personal reference scale: an approach to social judgment. In: Berkowitz, L. (Ed.), *Advances in Experimental Social Psychology* (Vol 4, pp. 315–370). [https://doi.org/10.1016/S0065-2601\(08\)60081-7](https://doi.org/10.1016/S0065-2601(08)60081-7).
- Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, 39(2), 111–133. https://doi.org/10.1207/s15326985ep3902_3
- VanderWeele, T. J. (2019). Principles of confounder selection. *European Journal of Epidemiology*, 34, 211–219. <https://doi.org/10.1007/s10654-019-00494-6>
- von Keyserlingk, L., Becker, M., Jansen, M., & Maaz, K. (2020). Effects of student composition in school on young adults' educational pathways. *Journal of Educational Psychology*, 112(6), 1261–1272. <https://doi.org/10.1037/edu0000411>
- Wouters, S., Germeijs, V., Colpin, H., & Verschueren, K. (2011). Academic self-concept in high school: Predictors and effects on adjustment in higher education. *Scandinavian Journal of Psychology*, 52(6), 586–594. <https://doi.org/10.1111/j.1467-9450.2011.00905.x>
- Zell, E., & Alicke, M. D. (2010). The local dominance effect in self-evaluation: Evidence and explanations. *Personality and Social Psychology Review*, 14(4), 368–384. <https://doi.org/10.1177/1088868310366144>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Herbert W. Marsh¹ · Reinhard Pekrun^{1,2,3} · Theresa Dicke¹ · Jiesi Guo¹ · Philip D. Parker¹ · Geetanjali Basarkod¹

Reinhard Pekrun
Reinhard.Pekrun@acu.edu.au

Theresa Dicke
Theresa.Dicke@acu.edu.au

Jiesi Guo
Guo@acu.edu.au

Philip D. Parker
Philip.Parker@acu.edu.au

Geetanjali Basarkod
Geetanjali.Basarkod@acu.edu.au

¹ Institute for Positive Psychology and Education (IPPE), Australian Catholic University, North Sydney 2060, Australia

² Department of Psychology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK

³ Department of Psychology, Ludwig Maximilian University Munich, 80802 Munich, Germany