Academic self-concept formation and peer-group contagion : Development of the big-fish-little-pond effect in primary-school classrooms and peer groups

Koivuhovi, Satu, Marsh, Herbert W., Dicke, Theresa, Sahdra, Baljinder, Guo, Jiesi, Parker, Philip D. and Vainikainen, Mari-Pauliina

This MS is the final prepublication (open access) version of the published article:

Koivuhovi, S., Marsh, H. W., Dicke, T., Sahdra, B., Guo, J., Parker, P. D., & Vainikainen, M.-P. (2022). Academic Self-Concept Formation and Peer-Group Contagion:Development of the Big-Fish-Little-Pond Effect in Primary-School Classrooms and Peer Groups. Journal of Educational Psychology, 114(1), 198–213. https://doi.org/10.1037/edu0000554

Academic Self-Concept Formation and Peer-Group Contagion: Development of the Big-Fish-Little-Pond Effect in Primary-School Classrooms and Peer Groups

How do peer groups influence academic self-concept formation? We evaluate developmental issues in the big-fish-little-pond effect (BFLPE; negative effects of class-average achievement on math self-concept [MSC]) and its generalizability to peer-group-average achievement (1,017 primary-school students tested in Years 4 and 6, 46 classes, 130 peer groups). The effects of peer-group-average and class-average achievement on MSC were both negative when we considered these two contextual effects separately. However, the effect of peer-group-average became nonsignificant in models with both contextual effects; the negative effect of class-average achievement was relatively unaffected. Results for peer-group-average achievement contradict predictions based on local dominance theory (that the BFLPE should be more negative for peer-group-average achievement than the more local frame of reference, a contrast effect) and social comparison choice studies (that peer-group-average achievement effects should be positive, an assimilation effect). Unsurprisingly, we found BFLPEs based on class-average achievement and gender differences favoring boys in both Years 4 and 6. However, consistent with theories of the cognitive development of social comparison and gender socialization/intensification processes, we also found negative effects of class-average and gender differences favoring boys on change in MSC (MSC in Year 6 controlling for MSC in Year 4) over this critical late-childhood period. Our results support the robustness of the BFLPE based on class-average achievement and developmental processes underpinning it, but do not support the posited effects of peer-group-average achievement.

Imagine a young girl, Kate, who is 10 years old (in Year 4 in primary school). Objectively, she is above average in terms of math achievement and a member of a peer group, friends who are also above average in math achievement. However, she is still developing her self-beliefs about her competencies in different school subjects— her academic self-concept (ASC) —that are likely to have long-term effects on her academic choices and accomplishments. Our research question is whether Kate will form her ASC based on her academic performance compared to her classmates as a whole or to a smaller number of classmates who form her peer group. Further, we ask whether being a member of an above-average peer group will have a positive or negative effect on her ASC and how this difference will change over the last 2 years of primary school. Although these questions have important implications, as detailed below, different intuitive perspectives and competing theoretical models provide conflicting answers to them. A peer group is a primary group of individuals who mutually choose to interact with each other based on links of friendship and common interests and who choose to work and play with each other. Peer-group members typically share similar characteristics (e.g., context, shared interest, age, geographical proximity, race, abilities, socioeconomic status; Brown & Larson, 2009; Cairns & Cairns, 1994; Hartup, 2005; Kindermann, 2016; Ryan, 2000; Wentzel & Caldwell, 1997). Peer groups are widely posited to provide a context for cognitive, social, and emotional development. They serve as a socialization agent in that their influence on thinking, identity formation, self-beliefs, and actions increases in importance during preadolescence and early adolescence (Gavin & Furman, 1989; Kindermann, 2007, 2016; Steinberg, 2010). Gest et al. (2008) noted that, ironically, most research on the effects of peers has focused on social outcomes even though the research is typically conducted in schools. In their research, Gest et al. (2008) found that by late primary school, academic reputations among a child's peers are well-formed and reciprocally related to subsequent academic self-concept and academic achievement. From these findings, it seems that the perspectives of one's peers influence one's subsequent ASCs and achievement, beyond the effects of prior achievement.

Molloy, Gest, and Rulison (2011; also see Kelley, 1952) distinguish between socialization and social comparison processes associated with peer groups. In directions for further research, Gest et al. (2008; Molloy et al., 2011) noted that peers provide an alternative frame of reference that children can use when making social comparisons that may influence their academic self-concepts—a major focus of the present investigation. However, the peer-group-average achievement will typically be correlated substantially with class-average achievement. Nevertheless, previous studies of peer-group contextual effects have largely neglected the fact that class-average achievement and related effects are confounded with the effect of peer-group-average achievement (i.e., on average, the peer-group-average achievement will be higher in classes where class-average achievement is higher). Thus, it is essential to juxtapose peer-group-average and class-average achievement and to disentangle the effects of these two contextual effects. Hence, this is a potentially important gap in peer-group research that we address.

We hypothesize peer groups to influence many academic and nonacademic outcomes, but unconfounding these peer-group effects from those of selection effects, competing contextual effects, and other preexisting differences is complex (Kindermann, 2007, 2016). As noted by Ryan (2000), the evaluation of the effects of peer groups is a multilevel issue, juxtaposing the effects of individual students and the peer groups to which they belong. Our focus here is on the influence of peer groups in late-primary-school years on math self-concept (MSC)—students' self-perceptions of their academic abilities and competencies in mathematics (Marsh, 2007). More specifically, using multilevel contextual models, we evaluate the effect of the peer-group-average achievement on MSC in Year 4 and Year 6 and juxtapose it with the effect of class-average achievement that has been the traditional approach in most previous research reviewed below. In doing so, we also unconfound the effect of peer groups within classes from the effects of classes in which the peer groups reside—an issue that has been largely ignored in peer-group research.

ASC, Social Comparison, and the Big-Fish-Little-Pond Effect

Psychologists from the time of William James (1890/1963) have recognized that self-concept is based on objective accomplishments evaluated in relation to frames of reference. This perspective is shared by many theoretical models based on, for example, relative deprivation (Davis, 1966), adaptation level (Helson, 1964), or social comparison processes (e.g., Festinger, 1954). In educational settings, the big-fish-little-pond effect (BFLPE; Marsh et al., 1984; Marsh & Parker, 1984) is the negative effect of school/class-average achievement on ASC. In other words, BFLPE predicts that a student in a class of high-achieving classmates will have a lower ASC than an equally able student in a class of average-achieving classmates.

Based on their review of 30 years of BFLPE research, Marsh and Seaton (2015; also see Marsh et al., 2014, 2017, 2018) claimed that the BFLPE was one of education and psychology's most cross-national, universal findings. Particularly strong support for this claim is based on multiple waves of nationally representative samples of 15-year-olds collected by the Organization for Economic Cooperation and Development (OECD) Program for International Student Assessment (PISA). Based on four cycles of PISA data (2000 –2012; *N* = 1,251,728), the BFLPE was negative in all but 1 of the 191 country samples and significantly so in 181 (Marsh, Parker, et al., 2019; Marsh & Seaton, 2015). In a recent meta-analysis of 33 BFLPE studies, Fang et al. (2018) reported that the negative effect of school-average ability was -.28 but that the effects tended to be somewhat smaller for primary-school students (-.21).

An extensive body of research shows that the BFLPE generalizes over a substantial number of individual student characteristics hypothesized as potential moderators of the BFLPE, including prior achievement, gender, goal theory constructs, implicit theories of ability, self-regulated learning strategies, and social interdependence measures (Marsh & Seaton, 2015; Marsh, Zanden, et al., 2019). BFLPE research provides robust support for the negative effects of class- and school-average achievement, the original theoretical model underpinning the BFLPE.

Local Dominance Effects

Most BFLPE studies (e.g., Guo et al., 2018; Marsh et al., 2015) are based on multilevel regression models with two levels; ASC is regressed on individual achievement and school- or class-average achievement. The effect of individual achievement on ASC is positive, but the effect of school- or class-average achievement is negative (the BFLPE). However, according to local dominance theory and research, the more local the frame of reference (i.e., the classroom rather than the school), the more important it tends to be (see Zell & Alicke, 2010) in the formation of self-evaluations. In support of the local dominance theory, Zell and Alicke (2010) conducted laboratory studies with random assignment to groups and experimentally manipulated feedback. They concluded that "when multiple comparison standards are available for self-evaluation, people rely on the most local comparison information while deemphasizing more general, and typically more diagnostic, forms of comparison feedback" (Zell & Alicke, 2010, p. 369). Integrating local dominance theory and BFLPE research, Marsh et al. (2014; also see Liem et al., 2013) used a three-level model to compare the relative importance of school-average and class-average achievement in a large, nationally representative sample of Dutch schools where there was extensive tracking between schools as well as ability tracks within schools. Consistent with local dominance theory (see Zell &

Alicke, 2010), Marsh et al. (2014) predicted and found that the negative effect of school-average ability was largely absorbed into the even more negative effect of class-average ability when both group-average effects were considered in the same model. Indeed, even though students were able to rank themselves with reasonable accuracy within classes, schools, and the country, ASC was almost exclusively determined in relation to class-average achievement—the most local context. However, the whole class might not be the most local frame of reference in that students within classes typically work and socialize within smaller peer groups. Here, we evaluate the generalizability of the BFLPE in relation to competing predictions about the effect of peer-group-average achievement in terms of its size, direction, and juxtaposition with the effect of class-average achievement on ASC. To the extent that the peer group provides a more local frame of reference than the class, the local dominance theory predicts that the effect of the peer-group-average achievement on academic self-concept would be negative and more negative than the effect of class-average achievement. Furthermore, when both class-average and peer-group-average achievement are considered in the same model, the negative effects of class-average achievement would be substantially attenuated and largely absorbed into the negative effects of peer-group-average achievement. There have apparently been no previous tests of local dominance theory in relation to BFLPEs based on class-average achievement and peer-group-average achievement. Here, we test these two contextual effects in a longitudinal study over the last 2 years of primary school (Years 4 – 6), the critical period of late childhood.

Assimilation and Explicit Social Comparison Effects With Individual Targets

Marsh (1984, 1987, 2007) emphasized that the contrast effect consistently observed in BFLPE studies is not the only potential outcome. In particular, being an average-ability student in a high-ability group of classmates may affect ASC such that it could be (a) below average because the frame of reference is established by the performance of above-average students (i.e., the BFLPE, a contrast effect), (b) above average as a consequence of membership in the high-ability grouping (i.e., a reflected glory or group identification, an assimilation effect), or (c) average because it is unaffected by the immediate context of the other students, or because (a) and (b) both occur and cancel each other. Thus, the observed negative BFLPE could be the combined effects of a large negative (contrast) effect and a smaller positive (assimilation) effect.

Although there is little support for assimilation effects in traditional BFLPE studies (Marsh & Seaton, 2015), other social comparison choice studies suggest that assimilation can occur in other situations. Indeed, the role that social comparison plays in the BFLPE has been the subject of extensive debate (e.g., Dai & Rinn, 2008; Marsh, Seaton, et al., 2008) and subsequent research attempting to integrate these theoretical perspectives and apparently conflicting results (see review by Marsh & Seaton, 2015). In BFLPE studies, the frame of reference is posited to be based on a generalized other, operationalized as the school- or class-average achievement. The comparison process is implicit, as students are not explicitly instructed to make comparisons with other individual students or a generalized other. However, there is also a considerable body of social psychological research into social comparison theory that focuses on specific individual comparison target persons explicitly chosen by an individual (Suls & Wheeler, 2000). Unlike BFLPE studies, the process is explicit in that participants are specifically instructed to choose one or a small number of comparison target persons. An upward comparison with a target person who is more able is posited to be a double-edged sword in that upward comparisons can be inspiring (assimilation effects) or deflating (contrast effects; Diener & Fujita, 1997; Major et al., 1991).

Relatedly, Akerlof and Kranton (2010) argue that group membership can lead to assimilation or reflected glory effects such that self-concept is positively related to the status of the group. Similarly, the identification-contrast model (Buunk & Ybema, 1997) posits that identification with a target can lead to assimilation. In contrast, Mussweiler et al. (2000) argue that assimilation is more likely when the person and the target are similar. Alternatively, Kelley (1952) posited that the direction of the effect is substantially determined more by the nature of the construct such that normative forces more influence identity formation, and comparative forces more influence self-appraisals.

Diener and Fujita (1997) conducted a comprehensive review of research from the perspective of social comparison choice research. They concluded that BFLPE studies provided the clearest support for predictions based on social comparison theory in what they referred to as an imposed social comparison paradigm. They noted that because "schools more nearly approximate 'total environments' (strongly controlling the information that the individual receives about the distribution of academic abilities, and also emphasizing feedback that has primarily a comparative meaning) than the environments examined in other comparison studies" (Diener & Fujita, 1997, p. 351). This imposed context can be distinguished for social comparison choice situations and peer groups where the context is typically chosen rather than imposed. In relation to chosen comparison targets, several school-based studies (Blanton et al., 1999; Huguet et al., 2001) reported that choosing a comparison target who is slightly more able than the self was associated with higher subsequent academic performance (see Marsh et al., 2010, for an alternative interpretation; also see Dicke et al., 2018) but had little effect on academic self-evaluations. Altermatt and Pomerantz (2005) more specifically evaluated the role of social comparisons for children's self-concepts within friendships. They found that when the reciprocal friend had higher achievement, children showed higher levels of achievement over time but also lower levels of academic self-concept (a contrast effect). However, Altermatt (2011) also noted social comparison processes through which friends benefit by sharing each other's successes (assimilation).

In an attempt to integrate social comparison choice and BFLPE studies, Seaton et al. (2008) collaborated with all the authors of the Blanton et al. (1999) and Huguet et al. (2001) studies in a reanalysis of results from both these studies. The selection of upward comparison targets did not affect self-evaluation in the reanalysis of the data by Blanton et al. but had some small positive effects for the data by Huguet et al. (2001). However, consistent with BFLPE studies, school-average ability had a negative effect on self-evaluations in the reanalysis of both studies. This led these authors to surmise that, perhaps, the positive effects of selecting individual upward comparisons coexisted with the negative effects on self-evaluations of implicit comparisons with the class-average achievement—the BFLPE.

In subsequent research, Huguet et al. (2009; also see Marsh, Trautwein, et al., 2008) juxtaposed the effects of class-average achievement and achievement levels of the comparison classmate chosen by the student. They found that class-average achievement was negatively related to academic self-concept (the contrast effect in BFLPE studies) but that the ability levels of classmates chosen as a target of comparison were positively related to academic self-concept (an assimilation effect). However, following Marsh, Trautwein, et al. (2008), Huguet et al. also asked each student whether the classmate he or she chose as a target was more or less able than the student making the choice. When they included this subjective comparison rating in the analyses, the effect of upward comparison was significantly negative (a contrast effect as in BFLPE studies). As proposed by Marsh, Trautwein, et al., these results suggest that apparent assimilation effects based on objective measures used in earlier social comparison choice studies may not generalize to the subjective differences actually perceived by students (also see Marsh et al., 2014; Parker, Marsh, et al., 2018). In the present investigation, our focus is on peer groups rather than explicitly chosen targets of social comparison. However, in many ways, the peer groups are more like the explicitly chosen target persons in social comparison choice studies that sometimes result in assimilation rather than the implicit, imposed comparisons based on the class-average achievement in BFLPE studies. From this theoretical perspective, the frame of reference based on the peer group might not result in the contrast effect as in traditional BFLPE studies. Indeed, it might even result in small assimilation effects that are consistent with reflective-glory effects associated with peer-group membership and assimilation effects found in social comparison choice studies. The direction of these assimilation effects would, of course, be exactly opposite to those predicted by local dominance theory and the BFLPE. Thus, our research is critical in determining whether the self-selected social comparison target paradigm or the local dominance paradigm— or neither—applies to contextual effects based on peer groups.

Developmental Issues in Social Comparison Processes and Gender Differences

Development of Social Comparison Processes

Social comparison processes underpinning our predictions about the contrast effects with the BFLPE and peer groups are hypothesized to be age related (e.g., Marsh et al., 2015). Most BFLPE research is based on high school students, but a growing body of research demonstrates the BFLPE with primary-school students (Dicke et al., 2018; Televantou, 2014; Tymms, 2001). Indeed, Tymms (2001; also see Televantou, 2014) found support for small but significant BFLPEs in U.K. students as young as 7 years of age. Based on cross-national data (Trends in

Mathematics and Science Study) from nationally representative samples of students in Years 4 and 8, Marsh et al. (2015) demonstrated the BFLPE was evident in both age groups but was stronger for older students. Marsh et al. speculated that this age difference was due to developmental differences in the cognitive ability of young children to formulate ASCs in relation to their own levels of achievement and those of their classmates. Similarly, Harter (2015; also see Marsh, 2007) notes that the emergence in middle childhood of the ability to use social comparisons for purposes of self-evaluation results in more accurate but also more negative self-appraisals as children mature. Although longitudinal studies of the BFLPE during primary school are rare, at least some research (Televantou, 2014; also see Dicke et al., 2018) suggests that the size of the effects grows larger as students develop the cognitive skills to incorporate social comparison processes into the formation of their academic self-concepts. We assume that these age-related differences in the BFLPE would also generalize to the effect of peer groups. However, as noted above, even the predicted direction of effects based on peer groups is less clear than in the BFLPE studies based on class/school-average achievement, and we know of no research explicitly pursuing this issue.

The BFLPE tends to be larger for older students. However, in an evaluation of age-related differences in the size of the BFLPE, Parker et al. (2019; also see Parker, Marsh, et al., 2018) subsequently emphasized that age is typically confounded with the extent of ability stratification in school systems. In other words, older students are more likely to be taught in schools and classes based on their ability and prior levels of academic achievement.

Consistent with the theoretical (social comparison) model underpinning the BFLPE, the BFLPE is based substantially on the extent of ability stratification in schools and classes. Thus, if the class-average achievement levels were similar in all classes within a given population, then the BFLPE would be predicted to be zero. On this basis, Parker et al. (2019) predicted that differences in the sizes of the BFLPE across OECD countries and over different age groups within the OECD countries could be explained in terms of the differences in ability stratification. Extending the rationale of this theoretical prediction, they suggested that the age-related differences in the sizes of the BFLPE found by Marsh et al. (2015) could be explained in terms of the greater prevalence of ability-grouping in secondary schools than primary schools, rather than developmental changes in the ability of young students to use social comparison processes.

In tests of these competing predictions, Parker et al. (2019) found that controlling ability stratification (intraclass correlations in relation to schools within each country) reduced but did not eliminate the age differences in the sizes of the BFLPE. Parker et al. thus concluded that the growth of the BFLPE with age was apparently a function of growth in cognitive maturity as well as ability stratification, suggesting that the use of social comparison processes is a capacity that is acquired as children develop and mature.

We note, however, that both the original Marsh et al. (2015) study and the subsequent Parker et al. (2019) study were based on cross-sectional data. Hence, there is a need for longitudinal comparisons that provide a more robust basis for evaluating developmental processes underpinning age-related differences found in cross-sectional studies. Furthermore, the cross-national comparisons used in both studies by Marsh et al. and Parker et al. potentially confounded country-specific characteristics, ability stratification, and age groups. In the present investigation, we pursue this issue by evaluating age-related changes in the size of the BFLPE with longitudinal data for primary-school students in Years 4 and 6. Because the composition of classes in our study was constant over this period, the results are not compromised by differences in ability stratification.

Development of Gender Differences

Gender differences in self-concept have long been a topic of interest to psychologists (see Wylie, 1979). Subsequently, researchers (e.g., Eccles, 1984; Eccles, Wigfield, et al., 1993; Marsh, 1989; Parker, Van Zanden, et al., 2018) reported that gender-stereotypic differences in multiple dimensions of self-concept (e.g., boys higher in MSC but lower in verbal self-concept) were consistent over the preadolescent to young-adult period. Across gender-stereotypic differences in multiple domains of self-concept, gender differences in MSC favoring boys were particularly substantial. In a review of gender difference research, Hyde (2005) proposed the gender similarity hypothesis, emphasizing that gender differences in most published research are very small and that men and women tend to be more alike than different. Nevertheless, in their classic cross-national study of gender differences based on the PISA data, Else-Quest et al. (2010) found only small differences in math achievement (effect size [ES] < .15) but much larger gender differences in favor of males for MSC (ES = .33). From a developmental perspective, Hill and Lynch (1983) suggested "gender-role intensification," in which conformity to gender role stereotypes becomes increasingly important with age. This is supported by reports that differences in the importance placed on mathematical and verbal competence between males and females grow larger with age (see Eccles, 1987; Wigfield et al., 1991). However, some longitudinal studies have shown consistent gender differences in math constructs over age (Fredricks & Eccles, 2002; Jacobs et al., 2002; Watt, 2004), whereas others suggest that the gender differences increase in size at least through early adolescence (Frenzel et al., 2010). Relatedly, Leaper (2013) emphasized that from an early age, children tend to form same-sex peer groups, which play an important role in shaping gender-typed norms. Furthermore, because peer groups for this age group are typically single sex, peer-group studies must include gender differences in that peer-group effects are likely to be confounded with gender differences.

The Present Investigation

The overarching aim of the present investigation is to evaluate the relative importance of the achievement levels in one's peer group compared to the class as a whole in the formation of academic self-concept. Integrating largely separate research literatures on the role of peer groups in the formation of academic self-concept and the BFLPE, we seek to extend and address gaps in both these research literatures.

The present investigation is part of a longitudinal study based on secondary data collected by the Centre for Educational Assessment at the University of Helsinki for primary-school students in Years 4 and 6 (1,017 students, 46 classes). As part of this data collection, students nominated up to five classmates who were part of their peer group that were used to identify 130 peer groups within the 46 classes (see Method for further discussion). Using longitudinal data, we test the BFLPE on MSC based on math test scores, class-average math test scores, and peer-group-average math test scores. We did separate tests for MSC in Years 4 and 6, as well as for change in MSC (Year 6 controlling MSC in Year 4 [MSC-Yr4]). Noting the need to control for preexisting differences, in subsequent analyses, covariates (teacher ratings of math skills, gender, parent education, and results from a school-readiness test from Year 1) were added to the model (Table 1). The school-readiness measure provided control for preexisting achievement, a desirable feature that is often not available in BFLPE studies. Based on the research reviewed earlier and the theoretical model underpinning the BFLPE, we offer the following research hypotheses:

Research Hypothesis 1: Consistent with BFLPE research, for both MSC-Yr4 and MSC in Year 6 (MSC-Yr6), the effect of student-level math test scores will be positive, and the effect of peer-group-average achievement will be negative (see Model 1A in Figure 1). Furthermore, change in MSC (MSC-Yr6 after controlling MSC-Yr4) will also be negatively related to peer-group-average achievement (see Model 1C in Table 2). We based this prediction on the finding that the BFLPE is a function of cognitive maturity such that social comparison is an age-related capacity that is acquired as children develop.

Research Hypothesis 2: Predictions based on peer-group-average achievement parallel those based on class-average achievement (Research Hypothesis 1). Consistent with BFLPE research, for both MSC-Yr4 and MSC-Yr6, the effect of student-level math test scores will be positive, and the effect of class-average math achievement will be negative (see Model 2A in Figure 1). Change in MSC (MSC-Yr6 after controlling MSC-Yr4) will also be negatively related to class-average achievement (see Model 2C in Table 2). This prediction is also based on the finding that the BFLPE is a function of cognitive maturity such that social comparison is a capacity that is acquired as children develop.

Research Hypothesis 3: Models of MSC-Yr4 and MSC-Yr6 contain both class-average and peer-group-average achievement (see Model 3A in Figure 1). For these models, local dominance theory predicts that the effect of class-average achievement will be largely absorbed into the effect of peer-group-average achievement and substantially smaller than the corresponding model that does not include peer-group-average achievement. This is a straightforward extension of the local dominance theory. We note, however, that alternative predictions exist based on social comparison choice studies when student choice is made explicit.

Research Hypothesis 4: The introduction of covariates (gender, parental education levels, a school-readiness measure from Year 1, and teacher ratings of achievement) allows us to control for potentially confounding correlates of MSC (see Model 4A in Figure 1). We hypothesize that their inclusion will not substantially change the pattern of results for the two contextual effects (class-average and peer-group-average achievement), but the gender differences and their change over the last 2 years of primary school might be substantively interesting in their own right over this developmentally important period.

Method

Procedure

The present investigation is part of a longitudinal study collected by the Centre for Educational Assessment at the University of Helsinki. Data consist of several measurement points where students' cognitive skills and self-beliefs were assessed. In addition, the data set includes sociometric tasks measuring students' peer relations. In autumn 2007, 17 schools were randomly selected from the schools in a large city in southern Finland using an equal-probability method that ensured representativeness with regard to socioeconomic status. In order to get a representative sample of students in the chosen city, the goal was to get 800 students into the sample. However, out of those 17 schools, one school with two regular classes declined to participate, making the final number of schools 16 at the beginning of the study. Also excluded were 19 small classes (with one to eight students in each) for children with very high special education needs or completely lacking the language skills to participate. This led to a sample with 16 schools and 40 classes. Parents were informed about the study through the Education Department of the city, securing the agreement of all the sampled pupils. From all families asked to participate in the study, two declined, and all others accepted to participate. Later, at the beginning of Grade 4 (autumn 2010), the sample size was increased, and four new schools with six classes were included in the sample as many students from the original sample had transferred to them.

Classroom teachers collected data during regular school hours for all but the sociometric data that were collected separately by the research coordinator. Data collections were at the beginning of the school year (from mid-August to mid-September 2010) for fourth grade and at the end of the school year (April 2013) for sixth grade. In fourth grade, the data questionnaire consisted of three parts. Each part was completed during one regular 45-min lesson (3 X 45 min). In sixth grade, all tasks were in one booklet, and completing it took time approximately 90 min (2 X 45 min). Students answered either in two different lessons or one double lesson.

Ethical and safety standards were met, and data were collected in accordance with American Psychological Association guidelines. Municipal school authorities reviewed the proposal and granted permission to collect the data. Families of all children were sent an information packet providing a summary of the project that was endorsed by the municipal school authorities. Parents completed a form agreeing to participate in the study. The anonymity of responses was maintained, and researchers had access only to deidentified data. Only data managers had access to a separate database where personal details were stored.

Participants

In the fourth-grade assessment in autumn 2010, there were 950 students present (53% girls; mean age M = 10.22 years, SD =.33), and in the sixth-grade assessment in spring 2013, there were 939 students (52% girls; mean age M = 12.81 years, SD = .33). Participants (N = 1,017) belonged to 130 peer groups within 46 classes.

Measures

We present means, standard deviations, and correlations among variables in Table 1. Math self-concept was based on responses to three items ("Math is very easy for me"; "I usually handle even the more difficult math problems well"; "I am good in math") using a 7-point Likert (1 =

not true at all to 7 = *very true*) response scale. Estimates of reliability were good for both Year 4 (cx = .86, w = .92) and Year 6 (cx = .91, w = .95).

School-Readiness Measure of Achievement

School-readiness measure of achievement was based on tests conducted in Year 1 (a brief test based on the ability to follow instructions, a visual-spatial memory task, and an analogic reasoning task; Hautamäki et al., 2001). The overall test consisted of 17 items (four following instructions, six visual-spatial memory items, and seven geometric analogy items). Each of the 17 items was scored dichotomously as correct or incorrect. The total score had moderate reliability (cx = .74, w = .59).

The Mathematics Achievement Test

The mathematics achievement test was measured with two tests measuring mathematical thinking skills: the Arithmetic subscale of the Wechsler Adult Intelligence Scale–Revised (WAIS-R: Wechsler, 1981) and the Arithmetic Operations test developed by Demetriou and his colleagues (Demetriou et al., 1991, 1996). Both tests measure the underlying competencies behind mathematics achievement, but they are not directly linked to curricular contents in mathematics (Vainikainen, 2014). The test included nine items (five from the Arithmetic subscale and four from the operations test). In the Arithmetic subscale of the WAIS-R (Wechsler, 1981), the teacher read aloud a mathematical problem (e.g., If you buy two bus tickets and one ticket costs 3 euros 50 cents, how much money do you get back if you give 10 euros?), and the pupils wrote down the answer in their test booklets. In Arithmetical Operations by Demetriou et al. (1991), a child had to solve arithmetical operations and complete hidden operators to equations in order to complete the equation. In each item, there were one to four hidden operators (e.g., [(5 a 3) b 4 = 6. In this task letter, a/b stands for: addition (+)/subtraction

(-)/multiplication (●)/division (+)?]). For present purposes, all nine items were first coded 0 –1 for a correct answer, and the math achievement score was the total number of correct responses across all items. The total score had reasonable reliability (cx = .65, w = .79).

Teacher Ratings of Achievement

Teacher ratings of achievement were collected at the fourth-grade assessment. Because students were not given traditional letter grades or numerical scores as a formal summary of their progress, teachers were asked for each student "to evaluate the pupil's current performance in math" (with school grades ranging from 4 to 10, where 4 stands for failed and 10 for excellent performance). Based on preliminary descriptive statistics, MSC is modestly and approximately equally correlated with math test scores (*r*s = .30 and .32 in Years 4 and 6; see Table 1) and teacher ratings of math achievement (*r* = .32 and .30 in Years 4 and 6).

Social Network Analysis

Students also completed a sociometric task in which they chose up to five other students from within their class with whom they (a) worked with on academic school tasks, (b) played with before and during school, and (c) played with after school. For present purposes, we focus on the frame of reference based on the academic accomplishments of classmates in one's peer groups in the formation of academic self-concept. Hence, we chose peer groups with whom students worked on academic tasks as being most relevant to the formation of ASC. Nevertheless, we did separate analyses for each of the three bases for peer-group formation. However, because the peer groups were nearly identical based on the three questions, the results based on the three different nominations were also nearly identical (see subsequent discussion). For simplicity of presentation, we based analyses reported on peer groups in relation to workgroups, but results for all three are presented, compared, and discussed.

To establish peer groups, we used a modified version of the procedure of Coie et al. (1982), which is widely used in research on peer-group contextual effects (e.g., see Parker et al., 2015; Sahdra et al., 2020). To extract peer group network communities within each class, we ran the InfoMap community detection algorithm on the peer network data using the R package *igraph* (Csardi & Nepusz, 2006). InfoMap is a highly efficient data compression approach to represent the community structure of a network. The details of the algorithm are described elsewhere (Rosvall et al., 2009; Rosvall & Bergstrom, 2007, 2008; also see Parker et al., 2015). Briefly, the goal of the algorithm is to minimize the information required to characterize a "random walk" through the network using a coding system that distinguishes communities in the network but also the nodes in any given community. The probability flow of random walks on a network is a proxy for information flows in real systems, and communities are detected by decomposing the network into modules by compressing a description of the probability flow. In a situation with sparse intercommunity links, for instance, InfoMap's walker tends to stay longer inside the communities, so only the community-level coding is needed to describe the path of the walker, thus representing the structure of communities. Directed links are considered in characterizing the walker's path. We focus on directed social networks (the fact that one individual cites another as a friend does not necessarily mean that the peer nomination is reciprocated). On the basis of these peer groups, peer-group-average levels of achievement were computed for each peer group.

Recent studies have shown that InfoMap, relative to many other algorithms, tends to represent the community structure that is the most similar to the known community structure of a reference network (Aldecoa & Marín, 2013; Orman et al., 2012). In tests employing the classic Lancichinetti–Fortunato–Radicchi benchmark graphs of known network properties, InfoMap consistently performs well. This is indicated by traditional test indices, such as the fraction of correctly classified nodes, rand index, adjusted rand index, and normalized mutual information, and in methods examining the topological properties of the estimated community structures (Orman et al., 2012). Lancichinetti–Fortunato–Radicchi models also show that InfoMap is especially accurate in community detection in networks with node sizes fewer than 1,000 (Yang et al., 2016). Even in highly stringent tests employing complex closed benchmarks, InfoMap tends to outperform other algorithms in detecting communities (Aldecoa & Marín, 2013). Like open benchmarks, closed benchmarks start with a network of known community structure, but unlike open benchmarks, they rewire the network such that the final network preserves the original network community structure but with nodes randomly assigned among communities. This process allows a quantitative evaluation of the optimal functioning of the algorithms used to detect known community structure (Aldecoa & Marín, 2013). No algorithm performs perfectly in such stringent tests, but InfoMap outperforms more than a dozen other contenders (Aldecoa & Marín, 2013). Therefore, InfoMap was deemed as the most suitable community detection algorithm for the purposes of this study.

Data Analysis

In the present investigation, the class composition of students in Years 4 and 6 was the same. Because of the nature of the data collection from intact classes, there were few missing data within waves. We tested a total of 1,017 students in either Year 4 or 6. Sample sizes varied somewhat according to the year and variable (see *N*s listed in Table 1). For example, MSCs were available for 945 students in Year 4 and for 881 students in Year 6; 833 students had MSCs for both years. These analyses were done with SPSS (Version 25). In order to accommodate missing data, we constructed multiple imputations based on all variables considered in the present investigation, including preliminary estimates of class- and peer-group average achievement scores; we constructed 50 imputed data sets.

Class-average and peer-group-average math test scores in Year 4, as well as the mean self-concept scale score for Years 4 and 6, were then recomputed using the imputed data. Here, as is typically the case (e.g., Pfaff, 2010), particularly for primary school, peer groups were nearly always gender specific. Based on this observation, for the 9% of students who were not assigned to a particular peer group (mostly due to missing data used to form peer groups), the within-class gender average test score was used as the peer- group-average test score. All variables except gender were standardized ($M = 0$, $SD = 1$) to facilitate interpretations of regression coefficients in relation to a standardized effect size. Based on the imputed data, a series of multilevel (three-level) regression models was conducted in which MSC at Year 4 or 6 was the dependent variable. Results from the 50 imputed data sets using Rubin's (1987) rules were based on the automated process available in SPSS. We conducted multilevel models in the following manner. We began by testing the most basic BFLPE model in relation to peer-group-average achievement (Model 1A in Table 2 and Figure 1), class-average achievement (as in most previous BFLPE studies; Model 2A in Table 2 and Figure 1), or both these contextual effects in the same model (Model 3A, Table 2 and Figure 1).

For Models 1A, 2A, and 3A, individual and aggregated math achievement were the predictor variables, and MSC-Yr4 was the dependent variable. For Models 1B, 2B, and 3B, the dependent variable was MSC-Yr6 (not controlling MSC-Yr4; i.e., the total effects of Year 4 predictors on MSC-Yr6). For Models 1C, 2C, and 3C, the dependent variable was the change in MSC (MSC-Yr6 after controlling MSC-Yr4; i.e., the direct effects of Year 4 predictors on MSC-Yr6 controlling MSC-Yr4).

Results

Class-Average Achievement and Peer-Group-Average Achievement (Research Hypotheses 1–3)

Our primary focus is juxtaposing contextual effects associated with peer-group-average achievement and class-average achievement (Table 2 and Figure 1), as well as predictions based on the BFLPE and local dominance theory (see Research Hypotheses 1–3 in the present investigation).

Research Hypothesis 1: Peer-Group-Average Achievement

We began with simple models (1A, 1B, and 1C) based on peer-group-average achievement. Consistent with Research Hypothesis 1, the effect of student-level math achievement on MSC was positive for both years (.39 on MSC-Yr4, Model 1A; .38 on MSC-Yr6, Model 1B), but the effect of peer-group-average achievement on MSC was negative for both years (-.14 on MSC-Yr4, Model 1A; -.08 on MSC-Yr6, Model 1B). However, in contradiction to Research Hypothesis 1, change in MSC (MSC-Yr6 controlling MSC-Yr4) was not statistically significantly related to peer-group-average achievement (-.02 on MSC-Yr6, Model 1C in Table 2).

Research Hypothesis 2: Class-Average Achievement

We next fit parallel models (2A, 2B, and 2C) based on class- average achievement. Consistent with Research Hypothesis 2, the effect of student-level math test scores on MSC was positive for both years (.39 for both MSC-Yr4 and MSC-Yr6). However, the effect of class-average achievement on MSC was negative in both years (-.21 and -.16, respectively). Also consistent with Research Hypothesis 2, change in MSC (MSC-Yr6 controlling MSC-Yr4) was negatively related to class-average achievement in Year 4 (-.11, Model 2C in Table 2). We note that in this respect, based on Model C, results based on class-average achievement differ from those based on peer-group-average achievement.

Research Hypothesis 3: Juxtaposing Peer-Group- Average and Class-Average Achievement

In tests of Research Hypothesis 3, we included both class-average-achievement and peer-group-average achievement in the same analysis (Models 3A, 3B, and 3C in Table 2; also see Model 3A in Figure 1). Particularly in relation to predictions based on local dominance theory and extensions of the BFLPE to peer groups, Model 3 is the most important. Model 3 is also more complex than either Model 1 or 2 because the two contextual effects are necessarily related (i.e., on average, the peer-group-average achievement will be higher in classes where class-average achievement is higher). However, the results were easily interpreted in that the effect of peer-group-average achievement was nonsignificant in all three models (3A, 3B, and 3C), while the effect of class-average-achievement was significantly negative in all three models. Furthermore, the negative effect of class-average achievement was nearly unaffected by the inclusion of peer-group-average achievement (marginally smaller in Model 3A than Model 1A, -.19 vs. -.21; marginally larger in Model 3B than Model 1B, -.20 vs. -.16; and marginally larger in Model 3C than Model 1A, -.16 vs. -.11). It is also important to emphasize that change in MSC for the late-preadolescent period (MSC-Yr6 controlling MSC-Yr4) was negatively related to class-average achievement (-.16, Model 3C) but was not significantly related to peer-group-average achievement.

Although we considered the effects of peer-group-average achievement in a number of different models in the subsequent discussion, we note that in none of these models was the effect of peer-group-average achievement statistically significant. These results provided a resounding rejection of our a priori prediction based on local dominance theory (Research Hypothesis 3). However, they provided no evidence of a positive effect of peer-group-average achievement (an assimilation effect) sometimes reported in social comparison choice studies.

Covariates and Developmental Issues Related to Self-Concept Formation (Research Hypothesis 4)

In Model 4, additional covariates were added to Model 3—teacher ratings of classroom achievement, gender, results of a school-readiness test from the start of school (Year 1), and parental education. The purpose of these covariates was to control for potentially confounding variables but also because of interest in the effects of these variables. Consistent with Research Hypothesis 4, their inclusion had only small effects on the critical contextual effects, class-average effect, and peer-group-average achievement. In Model 4C (compared to Models 2C and 3C), there was a small
drop in the size of the class-average achievement (e.g., -.13 in Model 4C compared to -.16 in Model 3C and -.11 in Model 2C; Table 2). The effect of peer-group-average achievement remained nonsignificant in Models 4A, 4B, and 4C, as was the case in Models 3A, 3B, and 3C.

As noted earlier, it was also of interest to note the size and direction of relations associated with these covariates (also see correlations in Table 1) in relation to MSC-Yr4, MSC-Yr6, and change over time (MSC-Yr6 controlling for MSC-Yr4). The effects of parental education and the prior measure of school-readiness had no significant effects (also see correlations in Table 1). Hence, we now briefly describe the effects of teacher ratings and gender differences that were significantly related to MSC in Model 4.

Teacher Ratings

Teacher ratings of student achievement had significant effects on MSC in both years (.26, MSC-Yr4, Model 4A; .23, MSC-Yr6, Model 4B). This was unsurprising as much previous research shows that teacher assessments (or school grades based on teacher assessments) are substantially related to MSC (Marsh, 2007; Marsh et al., 2014). However, it was interesting that teacher ratings in Year 4 also had a significant effect on change in MSC (MSC-Yr6 controlling MSC-Yr4) over the last 2 years of primary school. We also note that the effects of student achievement based on test scores showed a similar pattern of results. In Model 4, test scores positively affected MSC in both years (.29, MSC-Yr4, Model 4A;.27, MSC-Yr6, Model 4B) and for change MSC over this period (.19, Model 4C). Hence, both measures of achievement (test scores and teacher ratings) influenced MSC-Yr6 beyond their substantial effect on MSC-Yr4.

Gender Differences

Gender differences in favor of boys were substantial (.29, MSC-Yr4, Model 4A; .32, MSC-Yr6, Model 4B)— even after controlling for test scores, teacher ratings, and other covariates. Again, this was unsurprising as previous research consistently shows that girls have lower MSCs than boys (see earlier discussion). However, it was interesting that gender also had a significant effect on change in MSC (MSC-Yr6 controlling MSC-Yr4) over the last 2 years of primary school, indicating new gender differences in favor of boys in MSC-Yr6 beyond those in MSC-Yr4.

Discussion

Peer Groups and Local Dominance Effects

The major focus of the present investigation was on the role of peer groups in self-concept formation and the establishment of frames of reference that young students use to form their MSCs. Based on growing support for local dominance theory, we predicted and found that the effect of peer-group-average achievement would be negative (Research Hypothesis 1). We also predicted that this peer-group contextual effect would be more negative than the negative effect of class-average achievement and would largely absorb the negative effect of class-average achievement (Research Hypothesis 3). However, these a priori predictions based on Research

Hypothesis 3 were not supported. Although the effect of peer-group-average achievement on MSC was negative in both Years 4 and 6, this effect was explained entirely by the negative effects of class-average achievement. Hence, even the negative effect of peer-group-average achievement was apparently an artifact of the confounding of achievement levels in peer groups with the achievement levels in the classes in which the peer groups were situated.

Social Comparison: BFLPE Versus Social Comparison Choice Studies

In our review of relevant research on frame-of-reference effects, we noted previous studies pitting predictions based on the BFLPE and social comparison choice studies. Although both areas of research were based on social comparison effects, research predictions were apparently contradictory. Opposing the contrast effects (the negative effects of school/class average achievement) in BFLPE studies, there were mixed results in social comparison choice studies where students were instructed explicitly to select target classmates with whom to compare their results. Indeed, some of these choice studies showed assimilation effects in relation to the achievement levels of classmates chosen as targets of comparison. In a series of studies trying to bridge the gap between these two disparate sets of results, Huguet et al. (2009; also see Marsh et al., 2010; Marsh, Trautwein, et al., 2008) reported that class-average achievement was negatively related to achievement (the contrast effect in BFLPE studies). However, the actual ability levels of the student chosen as a target of comparison were positively related to achievement (an assimilation effect)—suggesting that contrast and assimilation effects can coexist. Nevertheless, when analyses in these social comparison choice studies were based on student subjective ratings of the ability levels of their comparison target students as opposed to objective ability levels based on test scores, there was support for contrast effects consistent with the BFLPE.

Our study, based on peer-group-average achievement (rather than subjective ratings of the ability levels of one's peers), seems to be similar to social comparison choice studies that sometimes resulted in assimilation. However, our study differs in that social comparisons were implicit, as in BFLPE studies, rather than explicit, as in social comparison choice studies where students were explicitly asked to choose someone with whom to compare their marks. Nevertheless, there was no evidence of any assimilation based on peer-group-average achievement test scores.

Notably, the focus of our study and BFLPE research is more generally on social comparison processes that are central to the formation of academic self-concept. However, Molloy et al. (2011; also see Kelley, 1952) distinguished between socialization and social comparison processes associated with peer groups. Social comparison processes for which the group average is a basis of comparison traditionally result in contrast effects that are evident in BFLPE studies. However, assimilation effects are more likely in relation to the formation of identity and values that are influenced by adherence to group norms. Although not a focus of the present investigation, it would be useful for future research to evaluate this distinction in relation to peer groups and BFLPE research more generally.

Competing Contextual Effects

A critical issue illustrated in our study is that it is important to look at competing contextual effects. If we had only considered the effects of peer groups in isolation (Model 1), we might have concluded that there are small, negative effects associated with peer-group-average achievement, a contrast effect consistent with other BFLPE studies. However, peer-group effects are typically confounded with other contextual effects. In the present investigation, the juxtaposition of results based on peer-group-average and class-average achievement (Model 3) demonstrated that peer-group-effects disappeared when class-average achievement was included in the model. In this sense, there was no unique effect of the peer group. Our interpretation of this finding is that the effect of peer-group-average achievement in Model 1 was an artifact of the confounding of the two contextual effects. In this application, the critical issue driving the BFLPE was the class-average achievement and not the peer-group-average achievement. Although not a focus of the present investigation, we also note that peer groups— particularly in primary schools—were almost entirely gender segregated. Thus, peer-group effects are likely to be confounded with gender differences. For this reason, it is also critical to control for gender differences in the evaluation of peer-group effects.

Lack of Support for Local Dominance Theory

What might be the reasons for the failure of the local dominance predictions? We note that there were small negative effects of peer-group-average achievement when considered in isolation. However, this was apparently an artifact of the confounding of peer-group-average and class-average achievement; on average, peer groups in classes with higher-achievement students necessarily had higher levels of achievement than peer groups in classes with lower-achieving students. Hence, to the extent that one's peer group is "more local" than one's class, our results violate predictions based on local dominance theory. There are, of course, differences between the contextual setting in our applied field study and experimental laboratory settings in most local dominance studies. The design of the laboratory studies makes explicit the performance levels of the other individuals in the more local group by reporting their experimentally manipulated results and how they compare with those given the participant who is making self-evaluations. This apparently also increases the saliency of the local group as a basis of comparison, while less locally dominant groups in these laboratory studies are not based on anyone actually seen or experienced by the experimental participants. In our applied field study in actual primary-school settings, students have ongoing regular contact with both their peers in their peer group and the other students in their class. In this sense, there is a less clear demarcation between the peer group posited to be more local and classmates posited to be less local. Also, it is unclear how well students know the objective achievement results of peers in their peer group. This consideration was important in social comparison choice studies that found results differed when based on objective and subjective achievement levels of chosen comparison-target students. Finally, we also note that in relation to the nature of feedback given to students in classroom settings by teachers, there is more likely to be a focus on the class as a whole rather than peer groups so that the class-average achievement is a more natural basis of comparison.

Peer Groups Based on Working Together and Friendship

We note that our results might depend on the way that peer groups were formed. We chose to form peer groups in relation to classmates with whom students preferred to work on school tasks rather than friendship groups per se. We made this choice as this seemed most relevant to the focus of this study in terms of academic frames of reference. However, it is important to emphasize that students were also asked to nominate classmates in relation to two other criteria that were more friendship oriented. Nominations based on the three questions were nearly identical. Consequentially, the effects of peer-group-average achievement were highly similar for the different peer-group nominations. In particular, there were no significant effects of peer group based on any of the three sets of peer-group nominations (see Table 3). Nevertheless, it might be that for older students in high school settings that peer groups based on friendship and on working together on academic tasks might be more differentiated. Hence, this is an area for further research.

Developmental Processes

Development of Social Comparison Processes

There has been almost no developmental focus in BFLPE studies in primary schools based on longitudinal data. In crosssectional studies reviewed earlier, Marsh et al. (2015) attributed age-related differences in the size of the BFLPE to the age-related development of social comparison processes, whereas Parker et al. (2019) showed that some, but not all, of these differences could be explained by differences in ability stratification in primary and secondary schools. Our results support these claims that social comparison processes grow stronger as young students mature and provide more robust tests of these claims than previous cross-sectional studies. Importantly, our results are based on a true longitudinal analysis showing that class-average achievement has a significant effect on changes in MSC over the last 2 years of primary school (MSC-Yr6 controlling MSC-Yr4). Furthermore, our results are not confounded by differences in ability stratification because both the students and the composition of classes were the same in Years 4 and 6. In summary, our results provide stronger support for the conclusion that age-related differences in the size of the BFLPE are due to developmental differences in the growth of social comparison processes.

Development of Use of Objective Measures of Achievement on Self-Concept Formation

We also found that both teacher ratings and test scores from Year 4 had significant effects on MSC in Years 4 and 6 and also on the change in MSC (MSC-Yr6 after controlling MSC-Yr4) over this critical period of late childhood. This finding is consistent with the theoretical model positing that self-concept formation is a function of cognitive maturity such that the incorporation of objective sources of feedback is a capacity that is acquired as children develop.

Development Issues in Gender Stereotypes

Gender differences are important in our study for two reasons. First, in our study as in previous research, peer groups in primary school are almost entirely gender-specific. Hence, the effects of peer groups are substantially confounded with the effects of gender. Thus, the effects that reflected gender differences might be mistaken for peer-group effects. For this reason, it is important to control for gender differences when evaluating the effect of peer groups, particularly in primary-school settings. However, because there were no effects of peer groups, controlling for gender did not alter these contextual effects.

Second, the gender differences and how they change over this critical period in late childhood are substantively exciting questions in their own right. Our results provide a potentially significant contribution to the study of gender differences in MSC. We note that our data are based on responses by primary-school students in a country (Finland) well known for gender equality in relation to international measures of gender equality (but also see related research on the gender-gap paradox showing that gender differences tend to be larger in more gender-equal countries; Else-Quest et al., 2010; Marsh et al., 2020). Nevertheless, consistent with many previous studies, we found that there are gender differences in MSC that are substantially larger than the relatively small gender differences in math achievement (see Table 1). Cross-sectional studies of gender differences (e.g., Eccles, Wigfield, et al., 1993; Marsh, 1989) suggest that stereotypic gender differences increase during the preadolescent to adolescent period and persist into adulthood.

Although we based our research on a relatively narrower age span, our longitudinal data provide stronger tests of developmental processes underpinning age-related gender differences in MSCs. In particular, even after controlling for substantial gender differences in MSC-Yr4 and holding constant the composition of classrooms, there are new, additional gender differences MSC-Yr6 (.24, Model 4C) beyond those in MSC-Yr4. Indeed, the effect of gender on MSC in Model 4C is larger than the effect of either math test scores or teacher ratings of math skills and only slightly smaller than the effect of prior MSC-Yr4.

It is also interesting to note that these gender stereotypes are apparently not evident in teacher perspectives in that gender differences in teacher ratings of math classroom achievement favoring boys are small (r = .09, Table 1) and similar in size to those based on math test scores (r = .07). Hence, even in a country with substantial gender equality, there is clear evidence of dramatic growth in stereotypic gender differences during a relatively brief preadolescent period. This gender difference is beyond what can be explained in terms of small gender differences in objective math achievement and teacher assessments of math achievement.

Strengths, Limitations, and Directions for Further Research

A particular strength, but also a limitation of our study, was the focus on the last 2 years of primary school in a school system in which the class composition remained constant. In this situation, changes over time were not confounded with changes in class composition. Relatedly, as noted by Brown and Larson (2009) and others, peer groups are mostly confined to self-contained classrooms at the primary-school level. In our study, we capitalized on this situation by limiting students' choice of peers to those within the same classroom. This greatly simplified the task of identifying peer groups through traditional sociometric techniques but also the statistical models used to evaluate the effects and interpretation of the results. However, this is also a limitation because it precluded the inclusion of peers not within the same class. More important, it is not a reasonable assumption for secondary schools, where students typically have many different classes, each with possibly different compositions of students in each. Hence, it is important to test the generalizability of our results to secondary school settings, but this will involve more complicated approaches to the identification of peer groups (see Kindermann, 2007) and more complicated multilevel models to analyze the results.

Our main findings were that the peer-group average achievement had no effect beyond that of the corresponding class-average achievement and the generalizability of the BFLPE based on class-average achievement. We note, however, that it is important to test the generalizability of these results as there may be other constructs for which the peer group is more important than the class. In relation to this issue, Kelley (1952) and others have emphasized the distinction between self-appraisals (like MSC) that are more influenced by comparative forces that result in contrast effects and identity formation that is more influenced by normative processes that might result in assimilation effects. Our results are consistent with this distinction in that MSC is clearly a self-appraisal, and the results demonstrated contrast effects in the form of a BFLPE.

However, in a different research context where the focus is on normative processes that drive identity formation, researchers might find assimilation effects rather that contrast effects. It is conceivable that the peer-group effects would be stronger than those of the class. Indeed, Parker et al. (2019) and Goldthorpe (2007) suggested that assimilation models will be more likely when reference groups are cohesive, are historically situated in wider cultural groups, are evident early in development, and have strong normative power. In this respect, gender stereotypes (e.g., boys are good at math, girls are good at language) are well established, are taught from an early age, and impose strong normative pressure. On this basis, Parker et al. suggested that it is not surprising that gender has a strong assimilative effect on ASCs. Nevertheless, we note that peer groups based substantially on gender apparently did not have any assimilation effects. Although clearly beyond the scope of the present investigation, issues regarding what constructs and under what conditions assimilation might occur constitute an important direction for further research. From a developmental perspective, our study showed that changes in MSC over the last 2 years of primary school (MSC-Yr6 controlling for MSC-Yr4) were related to gender, age, and objective measures of achievement as predicted on the basis of underlying developmental processes. Of central importance, class-average achievement negatively affected change in MSC, consistent with a prediction that social comparison processes develop with age. Based on a similar developmental perspective, academic achievement from Year 4 (both test scores and teacher ratings) affected change in MSC. We also note that the mathematics achievement test was relatively short and may not have covered all the important components of math achievement. Although expedient in terms of brevity, it would be desirable to use a longer, more encompassing achievement test and, perhaps, one more closely aligned to the math curriculum.

Finally, there was also a substantial gender difference in change in MSC over the last 2 years of primary school. Although these effects are clearly age related and significant from a statistical perspective, there is a need for further research evaluating the nature of the developmental processes underlying each of them.

Conclusions

To conclude, our study was apparently the first study testing local dominance theory in relation to BFLPE based on class-average achievement and peer-group-average achievement. Our results gave no evidence for the local dominance hypothesis as the negative effect of peer-group-average achievement vanished when peer-group and class-average achievements were considered together in the same three-level model. Our longitudinal data set enabled us to study also the developmental aspects of the formation of MSC and BFLPE. Our results showed that the role of social comparisons increased as students grew older. In addition, we found that gender differences in favor of boys increased over the last 2 years of primary school.

References

Akerlof, G. A., & Kranton, R. E. (2010). *Identity economics: How our identities shape our work, wages, and well-being*. Princeton University Press.

Aldecoa, R., & Marín, I. (2013). Exploring the limits of community detection strategies in complex networks. *Scientific Reports*, *3*, 2216. https://doi.org/10.1038/srep02216

Altermatt, E. R. (2011). Capitalizing on academic success: Students' interactions with friends as predictors of school adjustment. *Journal of Early Adolescence*, *31*, 174 –203. https://doi.org/10.1177/0272431610379414

Altermatt, E. R., & Pomerantz, E. M. (2005). The implications of having high-achieving versus low-achieving friends: A longitudinal analysis. *Social Development*, *14*, 61– 81. https://doi.org/10.1111/j.1467-9507.2005.00291.x

Blanton, H., Buunk, B. P., Gibbons, F. X., & Kuyper, H. (1999). When better-than-others compare upward: Choice of comparison and comparative evaluation as independent predictors of academic performance. *Journal of Personality and Social Psychology*, *76*(3), 420 – 430. https:// doi.org/10.1037/0022-3514.76.3.420

Brown, B. B., & Larson, J. (2009). *Handbook of adolescent psychology: Vol. 2. Peer relationships in adolescence*. John Wiley & Sons. https:// doi.org/10.1002/9780470479193.adlpsy002004

Buunk, B. P., & Ybema, J. F. (1997). Social comparison and occupational stress: The identification-contrast model. In B. P. Buunk and F. X. Gibbons (Eds.), *Health, coping, and well-being: Perspectives from social comparison theory* (pp. 359 –388). Erlbaum.

Cairns, R. B., & Cairns, B. D. (1994). *Lifelines and risks: Pathways of youth in our time*. Cambridge University Press. https://doi.org/10.1207/ s15327884mca0403_12

Coie, J. D., Dodge, K. A., & Coppotelli, H. (1982). Dimensions and types of social status: A cross-age perspective. *Developmental Psychology*, *18*(4), 557–570. https://doi.org/10.1037/0012-1649.18.4.557

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, *1695*(5), 1–9. http:// igraph.org

Dai, D. Y., & Rinn, A. N. (2008). The big-fish-little-pond effect: What do we know and where do we go from here? *Educational Psychology Review*, *20*(3), 283e317. https://doi.org/10.1007/s10648-008-9071-x

Davis, J. A. (1966). The campus as a frog pond: An application of the theory of relative deprivation to career decisions of college men. *American Journal of Sociology*, *72*, 17–31. https://doi.org/10.1086/224257

Demetriou, A., Pachaury, A., Metallidou, Y., & Kazi, S. (1996). Universal and specificities in the structure and development of quantitative-relational thought: A cross-cultural study in Greece and India. *International Journal of Behavioral Development*, *19*, 255–290. https://doi.org/ 10.1177/016502549601900203

Demetriou, A., Platsidou, M., Efklides, A., Metallidou, Y., & Shayer, M. (1991). The development of quantitative-relational abilities from childhood to adolescence: Structure, scaling, and individual differences. *Learning and Instruction*, *1*, 19 – 43. https://doi.org/10.1016/0959-4752(91)90017-3

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 comparisons and subjective well-being. In B. P. Buunk & F. X. Gibbons (Eds.), *Health, coping, and well-being: Perspectives from social comparison theory* (pp. 329 –357). Erlbaum.

Eccles, J. (1984). Sex differences in achievement patterns. *Nebraska Symposium on Motivation*, *32*, 97–132.

Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly*, *11*(2), 135–172. https://doi.org/ 10.1111/j.1471-6402.1987.tb00781.x

Eccles, J. S., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self-and task perceptions during elementary school. *Child Development*, *64*, 830 – 847. https://doi.org/10.2307/1131221

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*, 103–127. https://doi.org/10.1037/a0018053

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 Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology*, *38*(4), 519 –533. https://doi.org/10.1037/0012-1649.38.4.519

Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on*

Adolescence, *20*(2), 507–537.

Gavin, L. A., & Furman, W. (1989). Age differences in adolescents' perceptions of their peer groups. *Developmental Psychology*, *25*(5), 827– 834. https://doi.org/10.1037/0012-1649.25.5.827

Gest, S. D., Rulison, K. L., Davidson, A. J., & Welsh, J. A. (2008). A reputation for success (or failure): The association of peer academic reputation with academic self-concept, effort, and performance across the upper elementary grades. *Developmental Psychology*, *44*, 625– 636. https://doi.org/10.1037/0012-1649.44.3.625

Goldthorpe, J. H. (2007). *On sociology*. Stanford University Press.

Guo, J., Marsh, H. W., Parker, P. D., & Dicke, T. (2018). Cross-cultural generalizability of social and dimensional comparison effects on reading, math, and science self-concepts for primary school students using the combined PIRLS and TIMSS data. *Learning and Instruction*, *58*, 210 –219. https://doi.org/10.1016/j.learninstruc.2018.07.007

Harter, S. (2015). *The construction of the self: Developmental and sociocultural foundations*. Guilford.

Hartup, W. W. (2005). Peer interaction: What causes what? *Journal of Abnormal Child Psychology*, *33*(3), 387–394. https://doi.org/10.1007/ s10802-005-3578-0

Hautamäki, J., Arinen, P., Hautamäki, A., Lehto, J., Lindblom, B., Kupiainen, S., Outinen, K., Pekuri, M., Reuhkala, M., & Scheinin, P. (2001). *Ensiaskeleet— oppimisen edellytykset: Luokanopettajille tarkoitetun seulan toimivuus Helsinki-aineiston perusteella* [First steps— Prerequisites of learning: The functionality of the screening test for class teachers based on Helsinki-data].

Helson, H. (1964). *Adaptation-level theory: An experimental and systematic approach to* *behavior*. Harper & Row.

Hill, J. P., & Lynch, M. E. (1983). The intensification of gender-related role expectations during early adolescence. In J. Brooks-Gunn and A. C. Petersen (Eds.), *Girls at puberty* (pp. 201–228). Springer.

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*, 671–710. https://doi.org/10.1037/a0015558

Huguet, P., Dumas, F., Monteil, J. M., & Genestoux, N. (2001). Social comparison choices in the classroom: Further evidence for students' upward comparison tendency and its beneficial impact on performance. *European Journal of Social Psychology*, *31*(5), 557–578. https://doi.org/10.1002/ejsp.81

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*, 581–592. https://doi.org/10.1037/0003-066X.60.6.581

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, *73*(2), 509 –527. https://doi.org/10.1111/1467-8624.00421

James, W. (1963). *The principles of psychology* (Vol. *2*). Holt. (Original work published 1890)

Kelley, H. H. (1952). Two functions of reference groups. *Readings in Social Psychology*, *2*, 410 – 414.

Kindermann, T. A. (2007). Effects of naturally existing peer groups on changes in academic engagement in a cohort of sixth graders. *Child Development*, *78*(4), 1186 –1203. https://doi.org/10.1111/j.1467-8624.2007.01060.x

Kindermann, T. A. (2016). Peer group influences on students' academic motivation. In K. R. Wentzel & G. B. Ramani (Eds.), *Handbook of social influences in school contexts* (pp. 31– 47). Routledge Handbooks Online.

Leaper, C. (2013). Gender development during childhood. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Vol. 2. Self and other* (p. 326). Oxford University Press.

Liem, G. A. D., Marsh, H. W., Martin, A. J., McInerney, D. M., & Yeung,

A. S. (2013). The big-fish-little-pond effect and a national policy of within-school ability streaming: Alternative frames of reference. *Amer-ican Educational Research Journal*, *50*(2), 326 –370. https://doi.org/10.3102/0002831212464511

Major, B., Testa, M., & Blysma, W. H. (1991). Responses to upward and downward social comparisons: The impact of esteem-relevance and perceived control. In J. Suls & T. A. Wills (Eds.), *Social comparison: Contemporary theory and research* (pp. 237–260). Erlbaum.

Marsh, H. W. (1984). Self-concept: The application of a frame-of- reference model to explain paradoxical results. *Australian Journal of Education*, *28*, 165–181. https://doi.org/10.1177/000494418402800207

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

Marsh, H. W. (2007). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology*. British Psychological Society.

Marsh, H. W., Abduljabbar, A. S., Morin, A. J. S., Parker, P., Abdelfattah, F., Nagengast, B., &

Abu-Hilal, M. M. (2015). The big-fish-little-pond effect: Generalizability of social comparison processes over two age cohorts from Western, Asian, and Middle Eastern Islamic countries. *Journal of Educational Psychology*, *107*, 258–271. https://doi.org/10.1037/a0037485

Marsh, H. W., Kuyper, H., Morin, A. J., 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., Martin, A. J., Yeung, A., & Craven, R. (2017). Competence self-perceptions. In A. J. Elliot, C. Dweck, & D. Yeager (Eds.), *Handbook of competence and motivation* (pp. 85–115). Guilford.

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. (2020). Illusory gender-equality paradox, math self-concept, and frame-of-reference effects: New integrative explanations for multiple paradoxes. *Journal of Personality and Social Psychology*. Advance online publication. https://doi.org/10.1037/pspp0000306

Marsh, H. W., Parker, P. D., & Pekrun, R. (2019). Three paradoxical effects on academic self-concept across countries, schools, and students: Frame-of-reference as a unifying theoretical explanation. *European Psychologist*, *24*, 231–242. https://doi.org/10.1027/1016-9040/a000332

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*(2), 263–280.

https://doi.org/10.1037/dev0000393

Marsh, H. W., & Seaton, M. (2015). The big-fish–little-pond effect, competence self-perceptions, and relativity: Substantive advances and methodological innovation. In A. J. Elliot (Ed.), *Advances in motivation science* (Vol. 2, pp. 127–184). Elsevier. https://doi.org/10.1016/bs.adms.2015.05.002

Marsh, H. W., Seaton, M., Kuyper, H., Dumas, F., Huguet, P., Régner, I., Bunk, A. P., Monteil, J.-M., & Gibbons, F. X. (2010). Phantom behavioral assimilation effects: Systematic biases in social comparison choice studies. *Journal of Personality*, *78*(2), 671–710. https://doi.org/10.1111/ j.1467-6494.2010.00630.x

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. https://doi.org/ 10.1007/s10648-008-9075-6

Marsh, H. W., Smith, I. D., & Barnes, J. (1984). Multidimensional self- concepts: Relationships with inferred academic self-concepts and academic achievement. *Australian Journal of Psychology*, *36*, 367–386. https://doi.org/10.1080/00049538408255318

Marsh, H. W., Trautwein, U., Lüdtke, O., & Köller, O. (2008). Social comparison and big-fish-little-pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology*, *100*(3), 510 –524. https://doi.org/10.1037/0022-0663.100.3.510

Marsh, H. W., Zanden, B. V., Parker, P. D., Guo, J., Conigrave, J., & Seaton, M. (2019). Young women face disadvantage to enrollment in university STEM coursework regardless of prior achievement and attitudes. *American Educational Research Journal*, *56*, 1629 –1680. https://

doi.org/10.3102/0002831218824111

Molloy, L. E., Gest, S. D., & Rulison, K. L. (2011). Peer influences on academic motivation: Exploring multiple methods of assessing youth's most "influential" peer relationships. *Journal of Early Adolescence*, *31*, 13– 40. https://doi.org/10.1177/0272431610384487

Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin*, *26*(9), 1142– 1150. https://doi.org/10.1177/01461672002611010

Orman, G. K., Labatut, V., & Cherifi, H. (2012). Comparative evaluation of community detection algorithms: A topological approach. *Journal of Statistical Mechanics*, *2012*(8), 1–16. https://doi.org/10.1088/1742-5468/2012/08/P08001

Parker, P. D., Ciarrochi, J., Heaven, P., Marshall, S., Sahdra, B., & Kiuru, N. (2015). Hope, friends, and subjective well-being: A social network approach to peer group contextual effects. *Child Development*, *86*(2), 642– 650. https://doi.org/10.1111/cdev.12308

Parker, P. D., Dicke, T., Guo, J., & Marsh, H. W. (2019). A macro context theory of academic self-concept: Ability stratification and the big-fishlittle-pond effect. *PsychArXiv*. https://doi.org/10.31234/osf.io/ bwy59

Parker, P. D., Marsh, H. W., Jerrim, J. P., Guo, J., & Dicke, T. (2018). Inequity and excellence in academic performance: Evidence from 27 countries. *American Educational Research Journal*, *55*(4), 836 – 858. https://doi.org/10.3102/0002831218760213

Parker, P. D., Van Zanden, B., & Parker, R. B. (2018). Girls get smart, boys get smug: Historical changes in gender differences in math, literacy, and academic social comparison and achievement. *Learning and Instruction*, *54*, 125–137. https://doi.org/10.1016/j.learninstruc.2017.09.002

Pfaff, N. (2010). Gender segregation in preadolescent peer groups as a matter of class: Results from two German studies. *Childhood: A Global Journal of Child Research*, *17*(1), 43– 60. https://doi.org/10.1177/0907568209351550

Rosvall, M., Axelsson, D., & Bergstrom, C. T. (2009). The map equation. *European Physical Journal Special Topics*, *178*(1), 13–23. https://doi.org/10.1140/epjst/e2010-01179-1

Rosvall, M., & Bergstrom, C. T. (2007). An information-theoretic framework for resolving community structure in complex networks. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(18), 7327–7331.https://doi.org/10.1073/pnas.0611034104

Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(4), 1118 –1123. https://doi.org/10.1073/pnas.0706851105

Rubin, D. B. (Ed.). (1987). *Multiple imputation for nonresponse in surveys*. Wiley. https://doi.org/10.1002/9780470316696

Ryan, A. M. (2000). Peer groups as a context for the socialization of adolescents' motivation, engagement, and achievement in school. *Educational Psychologist*, *35*(2), 101–111. https://doi.org/10.1207/ S15326985EP3502_4

Sahdra, B. K., Ciarrochi, J., Parker, P. D., Craven, R., Brockman, R., Devine, E. K., Conigrave, J., & Chang, D. F. (2020). Discrimination as a frame-of-reference effect in overlapping friendship communities of ethnically diverse youth. *Cultural Diversity and Ethnic Minority Psychology*, *26*(1), 71– 81. https://doi.org/10.1037/cdp0000247

Seaton, M., Marsh, H. W., Dumas, F., Huguet, P., Monteil, J. M., Régner, I., Blanton, H., Buunk, A. P., Gibbons, F. X., Kuyper, H., Suls, J., & Wheeler, L. (2008). In search of the big fish: Investigating the coexistence of the big-fish-little-pond effect with the positive effects of upward comparisons. *British Journal of Social Psychology*, *47*(1), 73–103. https://doi.org/10.1348/014466607X202309

Steinberg, L. (2010). *Adolescence*. McGraw-Hill.

Suls, J., & Wheeler, L. (Eds.). (2000). A selective history of classic and neo-social comparison theory. *Handbook of social comparison* (pp. 3–19). Springer. https://doi.org/10.1007/978-1-4615-4237-7_1

Televantou, I. (2014). *Addressing an old issue from a new methodological perspective: A proposition on how to deal with bias due to multilevel measurement error in the estimation of the effects of school composition* (Unpublished doctoral dissertation). Oxford University, Oxford.

Tymms, P. (2001). A test of the big fish in a little pond hypothesis: An investigation into the feelings of seven-year-old pupils in school. *School Effectiveness and School Improvement*, *12*, 161–181. https://doi.org/10.1076/sesi.12.2.161.3452

Vainikainen, M. P. (2014). *Finnish primary school pupils' performance in learning to learn assessments: A longitudinal perspective on educational equity* (Unpublished doctoral dissertation). University of Helsinki, Helsinki.

Watt, H. M. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th-through 11th-grade Australian students. *Child Development*, *75*(5), 1556 –1574. Wechsler, D. (1981). *WAIS-R: Manual: Wechsler Adult Intelligence Scale-Revised*. Harcourt Brace Jovanovich for Psychological Corp.

Wentzel, K. R., & Caldwell, K. (1997). Friendships, peer acceptance, and group membership: Relations to academic achievement in middle school. *Child Development*, *68*, 1198 –1209.

Wigfield, A., Eccles, J. S., Mac Iver, D., Reuman, D. A., & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology*, *27*(4), 552.

Wylie, R. C. (1979). *The self-concept* (Vol. *2*). University of Nebraska Press.

Yang, Z., Algesheimer, R., & Tessone, C. J. (2016). A comparative analysis of community detection algorithms on artificial networks. *Scientific Reports*, *6*, 1–16. https://doi.org/10.1038/srep30750

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

Table 1

Descriptive Statistics and Correlations Among Variables

Variable	*N*	1	2	3	4	5	6	7	8	9
1. Math self-concept Year 4	945	**1.00**								
2. Math self-concept Year 6	881	**.36**	**1.00**							
3. Math test Year 4	943	**.30**	**.32**	**1.00**						
4. Teacher rating of math achievement Year 4	945	**.32**	**.30**	**.42**	**1.00**					
5. Gender (1 = girl, 2 = boy)	1,004	**.18**	**.18**	**.07**	**.09**	**1.00**				
6. Highest parental educational level	824	−.02	**.09**	**.14**	**.11**	.02	**1.00**			
7. School readiness test Year 1	602	.07	**.09**	**.15**	**.12**	.02	−.03	**1.00**		
8. Class-average math test Year 4	1,017	−.02	.05	**.43**	**.14**	.01	**.23**	.02	**1.00**	
9. Peer-group-average math test Year 4	1,017	.05	**.11**	**.53**	**.21**	**.13**	**.23**	.03	**.80**	**1.00**
M		**5.00**	**4.58**	**3.30**	**8.08**	**1.47**	**3.97**	**9.00**	**3.29**	**3.27**
SD		**1.51**	**1.50**	**1.83**	**1.40**	**.50**	**1.44**	**5.14**	**.81**	**.91**
Range		**1–7**	**1–7**	**0–9**	**4–10**	**1–2**	**1–6**	**1–17**	**0–9**	**0–9**

Note. Correlations are based on imputed data (i.e., all $N = 1{,}017$). However, the sample size (N) is based on the number of cases prior to imputation. Statistically significant correlations ($p < .05$) are presented in bold.

Table 2

Contextual Effect Models Juxtaposing Class-Average Achievement and Peer-Group-Average Achievement

Variable	Model A (Year 4)		Model B (Year 6)		Model C (Year 6)	
	Est	*SE*	Est	*SE*	Est	*SE*
Model 1						
Intercept	.02	.05	.05	.04	.05	.04
Math test Year 4	**.39**	**.04**	**.38**	**.04**	**.25**	**.04**
Peer-group-average math test scores	***−.14***	***.04***	***−.08***	***.04***	*−.02*	*.04*
Math self-concept Year 4					**.33**	**.03**
Model 2						
Intercept	.03	.04	.06	.04	.05	.05
Math test Year 4	**.39**	**.03**	**.39**	**.04**	**.27**	**.04**
Class-average math test scores	***−.21***	***.04***	***−.16***	***.04***	***−.11***	***.04***
Math self-concept Year 4					**.32**	**.03**
Model 3						
Intercept	.02	.04	.06	.04	.05	.05
Math test year 4	**.39**	**.04**	**.38**	**.04**	**.25**	**.04**
Peer-group-average math test scores	*−.02*	*.05*	*.06*	*.06*	*.07*	*.05*
Class-average math test scores	***−.19***	***.06***	***−.20***	***.06***	***−.16***	***.06***
Math self-concept Year 4					**.32**	**.03**
Model 4						
Intercept	**−.42**	**.10**	**−.43**	**.11**	**−.31**	**.11**
Math test Year 4	**.29**	**.04**	**.27**	**.04**	**.19**	**.04**
Peer-group-average math test scores	*−.08*	*.05*	*.00*	*.06*	*.02*	*.05*
Class-average math test scores	***−.13***	***.06***	***−.16***	***.06***	***−.13***	***.06***
Teacher rating Year 4	**.26**	**.04**	**.23**	**.04**	**.16**	**.04**
Gender (1 = female, 2 = male)	**.29**	**.06**	**.32**	**.06**	**.24**	**.06**
Parental education	−.05	.04	.06	.04	.07	.04
School readiness Year 1	−.01	.05	.03	.06	.03	.06
Math self-concept Year 4					**.28**	**.04**

Note. We test contextual effects (marked in italics) in a set of 12 models. Model 1A is the basic big-fish-little-pond effect (BFLPE) model showing the positive effect of individual math test scores on math self-concept (MSC) in Year 4 and the negative effects of class-average math test scores. Model 2A is based on peer-average math test scores instead of class-average test scores. Model 3A includes both class-average and peer-group-average test scores in the same model. Model 4A adds teacher ratings of classroom achievement as well as gender, parental education, and school readiness at Year 1 as covariates. Models A–C, respectively, are based on MSC in Year 4 (A), MSC in Year 6 not controlling Year 4 MSC (B), and MSC in Year 6 controlling Year 4 MSC (C). Thus, Model B reflects the total effects of Year 4 predictors on MSC in Year 6, while Model C reflects the direct effects of Year 4 predictors on MSC in Year 6, controlling MSC in Year 4. Estimates are standardized regression effects to facilitate interpretation is a standard effect size metric. Statistically significant effects ($p < .05$) are presented in bold.

Table 3

Contextual Effect Models Juxtaposing Class-Average Achievement and Peer-Group-Average Achievement

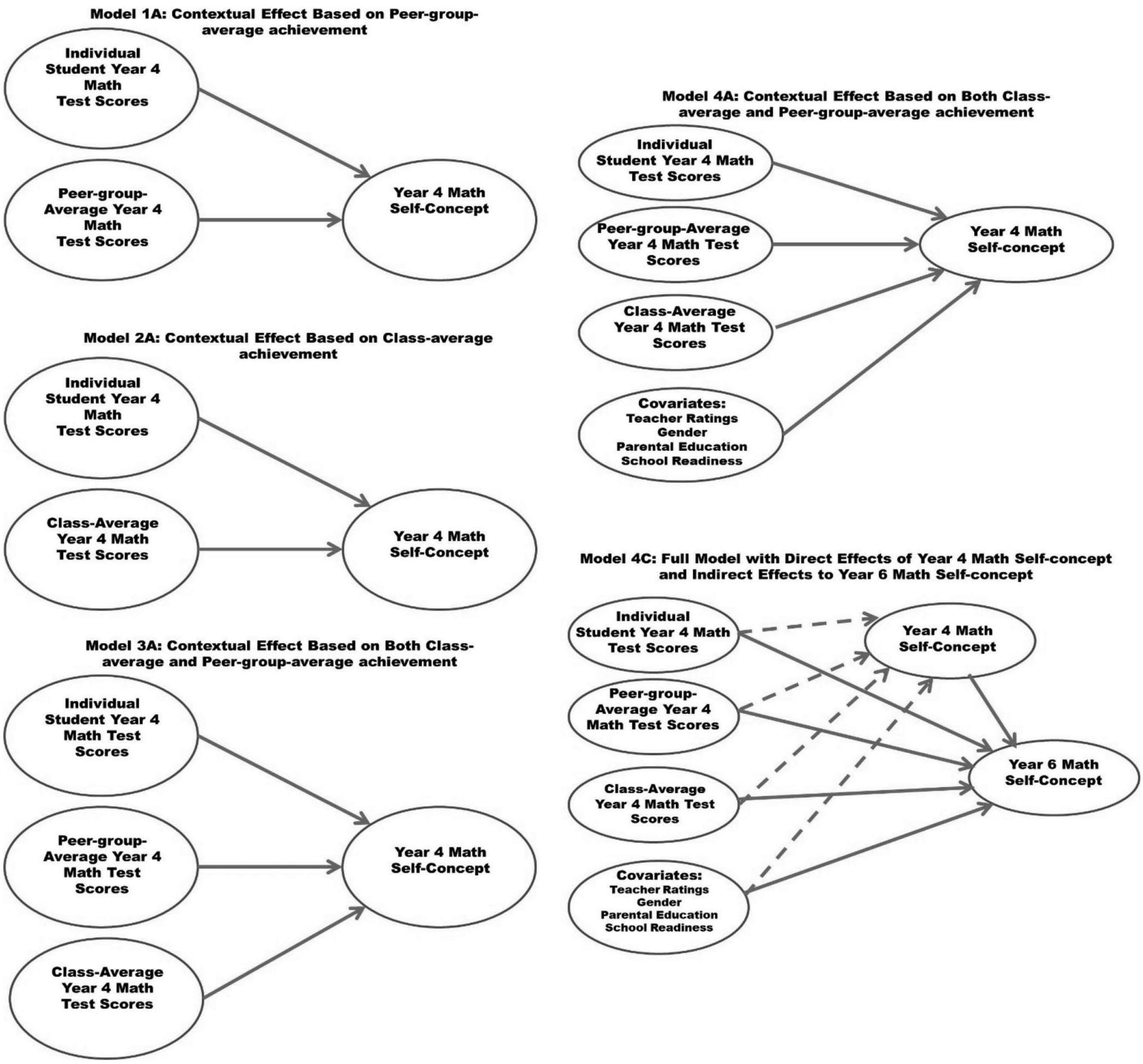
Model 4C	School work group		School play groups		After-school play groups	
Intercept	**−.31**	**.11**	**−.30**	**.11**	**−.29**	**.11**
Math test year 4	**.19**	**.04**	**.19**	**.04**	**.19**	**.04**
Peer-group-average math test Year 4	*.02*	*.05*	*.01*	*.05*	*.03*	*.05*
Class-average math test Year 4	***−.13***	***.06***	***−.12***	***.06***	***−.12***	***.06***
Teacher rating Year 4	**.16**	**.04**	**.16**	**.04**	**.16**	**.04**
Gender (1 = female, 2 = male)	**.24**	**.06**	**.23**	**.06**	**.22**	**.06**
Parental education	.07	.04	.07	.04	.07	.04
School readiness Year 1	.03	.06	.03	.06	.03	.06
Math self-concept Year 4	**.28**	**.04**	**.28**	**.04**	**.28**	**.04**

Note. We test contextual effects (marked in italics) in relation to three different questions used to form peer groups: who you work with on academic school tasks (Workgroups), who you play with before and during school (Play Before), and who you play with after school (Play After). All three analyses are based on Model 4C, the final model used in the set of analyses presented in Table 2. These results are for the effects of peer-group-average achievement and class-average achievement in math self-concept in Year 6, controlling for math self-concept in Year 4, math achievement (test scores and teacher ratings), gender, parental education, and a school readiness measure from Year 1. The results for peer groups based on workgroups are the same as presented in Table 2 of the main text but are presented again to facilitate comparisons. Estimates are standardized regression effects to facilitate interpretation is a standard effect size metric. Statistically significant effects ($p < .05$) are presented in bold. Results, particularly in relation to peer groups, are the same for all three sets of peer groups. This is not surprising in that peer groups based on the three different peer-group-nomination questions were nearly identical.

Figure 1

Path Diagrams Representing a Selected 5 of 15 Path Models Presented in Table 2

Designed to Test the Contextual



Effects of Class- Average Math Test Scores and Peer-Group-Average Achievement Math Test Scores

Note. Model 1A is the basic big-fish-little-pond effect (BFLPE) model showing the positive effect of individual math test scores on math self-concept (MSC) in Year 4 and the negative effects of class-average math test scores. Model 2A is based on peer-average math test scores instead of class-average test score. Model 3A includes both class-average and peer-group-average test scores in the same model. Model 4A includes four covariates (teacher ratings of class achievement, gender, parental education, and results of a school-readiness test administered in Year 1). In Model 4C, the dependent variable is MSC in Year 6, controlling MSC in Year 4; the dashed lines reflect the direct effects of the Year 4 predictor variables, controlling the effects of Year 4 MSC— change in MSC over the past 2 years of primary school associated with each of the predictor variables. Not depicted are the set of models (labeled 1B– 4B in Table 2) showing the effects of Year 4 predictor variables on MSC in Year 6 without controlling MSC in Year 4 —the total effects of the predictor variables, as well as additional Models 1C–3C showing change in MSC based on predictor variables in Models 1A–4C.