Phantom and big-fish-little-pond-effects on academic self-concept and academic achievement : Evidence from English early primary schools

Televantou, Ioulia, Marsh, Herbert W., Dicke, Theresa and Nicolaides, Christos

Televantou, I., Marsh, H. W., Dicke, T. and Nicolaides, C. (2021). Phantom and big-fish-little-pond-effects on academic self-concept and academic achievement : Evidence from English early primary schools. *Learning and Instruction*, *71*, Article 101399. https://doi.org/10.1016/j.learninstruc.2020.101399

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

Televantou, I., Marsh, H. W., Dicke, T., & Nicolaides, C. (2021). Phantom and big-fish-little-pond-effects on academic self-concept and academic achievement: Evidence from English early primary schools. Learning and Instruction, 71, 101399

https://doi.org/10.1016/j.learninstruc.2020.101399

Phantom and Big-Fish-Little-Pond-Effects on academic self-concept and academic achievement: Evidence from English early primary schools

Abstract

The Big-Fish-Little-Pond Effect (BFLPE) suggests that school-average achievement has a negative effect on academic self-concept (ASC); some research has also verified a negative effect on students' academic achievement. Our study evaluates the compositional effects of school-average achievement on both outcomes, using a longitudinal sample of English early primary school students in Year 1 and Year 4. We provide evidence for BFLPEs in children as young as six to nine years of age. Further, we show that the BFLPE is a potential mechanism in the negative compositional effect of school average achievement in Year 1 on students' achievement in Year 4. Once adjustments for measurement error are made, the negative effect of school-average achievement on students' self-concept, and on their subsequent achievement, increases. Our findings question previous research suggesting that attending a school with higher average achievement necessarily advances students' cognitive and affective outcomes.

1. Introduction

Defining self-concept broadly, Shavelson, Hubner and Stanton (1976) designate it a person's perception of self. Academic self-concept (ASC) then, refers to that specific component of self-concept that denotes the way in which individuals perceive their academic abilities and competencies in a specific subject (Byrne & Shavelson, 1986). ASC has been recognized for over fifty years as a fundamental responsibility of schools (Zirkel, 1971, p. 211). And, as much as ASC has been valued as an educational outcome in its own right, its relationship with other achievement results has also been widely addressed, since it has been shown to act as a mediator in the development of other desirable outcomes (Guay, Larose & Boivin, 2004; Marsh & Yeung, 1997). Importantly, there is solid evidence that academic self-concept and academic achievement are reciprocally related (REM: the reciprocal effects model; Marsh & Craven, 2005, 2006), so that higher ASC facilitates higher academic achievement, and vice versa.

An individual's ASC is particularly affected by environmental influences and by significant others – what the literature refers to as the reference group (Marsh et al., 2020; 2018; Pekrun et al., 2019). Specifically, in educational settings, since educational processes take place within classes or schools, students form their ASC by comparing their own accomplishments with those of their classmates or of the school at large. Hence, while a student with higher academic achievement will evidently have a higher ASC, the question arises as to what the impact of the school- /class-*average* achievement level is on students' ASCs.

1.1 Big-Fish-Little-Pond-Effects of Negative Compositional Effects

The Big-Little-Pond-Effect (BFLPE) hypothesis, the focus of the present study, predicts a negative effect of school (or class) average achievement on ASC, even though individual achievement is positively related to a student's ASC (Marsh, 1987; Marsh & Parker, 1984).

Thus, a student with a given achievement level is expected to have a lower ASC if they attend a school with higher average achievement, as opposed to how they would feel if they attended a school with a lower average achievement.

[*Insert Figure 1 here*]

The theoretical premise of the BFLPE lies in social comparison theory (Festinger, 1954; Aral & Nicolaides, 2017), and emphasises the need to consider the relative frames of reference in order to understand how people perceive their competencies in certain domains (Marsh, Kuyper, Morin, Parker & Seaton, 2014). The BFLPE is the net effect between *positive assimilation* (Marsh, Chessor, Craven & Roche, 1995; Marsh, Kong & Hau, 2000) due its affiliation to a prestigious institution or a highly selective educational program, and *negative contrast*. The latter, predicts that when students find themselves in high-achieving educational environments, due to comparison that they use to evaluate themselves, including comparisons with the achievements of their peers in their immediate environment – i.e., the school or classroom – they are predicted to have lower ASCs than if they were attending a low- or medium-achieving school.

Even though the theoretical foundations of BFLPE have mainly focused on ASC as the outcome variable, the policy implications of BFLPE may be relevant not only to research on ASC but also to other cognitive and affective outcomes. The BFLPE, together with the REM - the latter suggesting that mutually beneficial effects exist between academic achievement and academic self-concept (Marsh et al., 2005; see above) - predicts a negative effect of school average achievement on students' academic achievement (Dicke, Marsh, Parker, Pekrun, Guo, Televantou, 2018; Marsh 1987, 1991; Marsh & O' Mara, 2010). This contradicts a common assumption made by many parents in choosing a school for their children: namely, that the higher the achievement of the school, the better it is for the child's academic development. It is also at odds with the positive effect of school average achievement on students' academic progress that is commonly reported in educational effectiveness studies (the *peer spillover effect*; Fruehwirth, 2013; Willms, 1985).

1.2 Educational Effectiveness Findings of Positive Compositional Effects: Peer Spillover Effects

In the field of educational effectiveness (Creemers, Kyriakides & Sammons, 2010), class- and/or school-level aggregates of student achievement are used to evaluate the effects of the school's composition on student outcomes. When a student's performance in a school is affected by the characteristics of his or her fellow students, as quantified by the average achievement of their school peers (Marsh et al., 2000; Lüdtke, Marsh, Robitzsch, Trautwein, Asparouhov & Muthén, 2008), this gives rise to the predominance of *school compositional effects* (Perry, 2018; Televantou, Marsh, Kyriakides, Nagengast, Fletcher & Malmberg, 2015). Educational effectiveness research*, in contrast to BFLPE research and the reciprocal effects model*, generally implies a *positive* school compositional effect of average achievement – the *peer spillover effect*, suggesting that students achieve higher subsequent academic results than otherwise comparable pupils when they are instructed in schools (or classrooms) with lower achieving peers. Empirical evidence that supports this view can be found in the work of Hutchinson (1993), Willms (1985), Teddlie, Stringfield and Reynolds (1999), as well as in more recent studies (Stäbler, Dumont, Becker & Baumert, 2017).

While the prevalence of positive compositional effects seems intuitively reasonable, the reported positive compositional effects of achievement are often weak (Smith & Tomlinson, 1989; Gray, Jesson & Sime, 1990). Moreover, conflicting evidence has been shown in studies finding negative school compositional effects (Woodhouse, Yang, Goldstein & Rasbash, 1996; Tymms, 2001). Some educational effectiveness studies have queried whether school composition effects exist at all, as they have found little or no evidence of school compositional effects on achievement (Gibbons & Telhaj, 2012; Boonen, Speybroeck, de Bilde, Lamote, Van Damme & Onghena, 2014; Lavy, Silva & Weinhardt, 2012; Marks, 2015).

The lack of consensus in inferences on the magnitude and direction of the school composition effect in educational effectiveness studies has been attributed in part to the prevalence of student-level measurement error in the underlying data (Hutchison, 2007; Perry, 2018; Pokropek, 2015).

Specifically, a positive bias has been shown (Harker & Tymss, 2004; Televantou et al., 2015) so that educational effectiveness findings of positive school compositional effects of achievement (peer spillover effects) are, in fact, spurious.

1.3 Measurement Error as a Source of Bias

Measurement error may result in serious biases in estimating the effect of school-average achievement on a student's outcome, academic achievement or self-concept: this is something that has been demonstrated both mathematically (Marsh et al., 2009, 2012) and empirically (Gray et al., 1990; Hutchinson, 2004, 2007; Marsh et al., 2010; Woodhouse et al., 1996). Harker and Tymms (2004) coined the phrase 'phantom effects' on the basis of their finding that positive effects only appeared when measurement error was added to the data — now you see it, now you do not. When measurement error bias is corrected for, they become less-positive, non-significant, or even negative. Against this background, Dicke et al. (2018) claim, and, empirically show, that correcting for measurement error may be the key to achieving convergence between BFLPE research findings of negative compositional effects on self-concept and the educational effectiveness research findings of positive compositional effects on achievement. The methodology employed by Dicke et al.'s study – and ours – is described immediately below.

1.4 Statistical Modelling of Compositional Effects: Corrections for Measurement Error Bias

When the interest lies in measuring the school-level aggregate effects of achievement on individual outcomes, multilevel modelling (Snijders & Bosker, 2012) is incorporated, since this approach accommodates the nesting of the data (e.g. students nested in schools). Student-level is the lower unit of the analysis; students can also be referred to as individual-level or level 1 (L1-) units of analysis. Similarly, school-level often refers to the higher-level or level 2 (L2-) unit. When the research focus is on the evaluation of the BFLPE, the student-level outcome (ASC) is regressed on average achievement at the school-level, and on individual achievement at the student-level. The estimated effect of the aggregated variable over and above that of the corresponding individual-level characteristic on the outcome of interest is the *compositional* effect (Nash, 2003).

Increasingly sophisticated multilevel *latent* models have been developed that allow for adjustments of measurement error in the estimation of compositional effects (Lüdtke et al., 2008; Marsh et al., 2009). Surprisingly, only a few studies (Dicke et al., 2018; Nagenagast & Marsh, 2012; Televantou et al., 2015) have made use of this methodology. For instance, Marsh et al. (2009) distinguish between the *doubly manifest* approach, which is the conventional multilevel modelling approach to compositional analysis, and *multilevel latent variable* models (Table 1), which are capable of handling *measurement error* in individual- and school-level measures, either with (the *doubly latent* approach) or without (the *latent manifest* approach) corrections for *sampling error* in higher-level aggregates (school-average achievement).

1.5 The Present Study

In the present investigation, we sought to verify the BFLPE for students in the early stages of primary schooling. Our focus was on juxtaposing the BFLPE estimates obtained using models that control for measurement error bias, with models that do not. To this end, we used a large sample of English primary students. Year 1 and Year 4 mathematics achievement and self-concept (SC) measures were employed; the data sources were obtained from the Performance Indicators at Primary School (PIPS) project (Tymms, Jones, Albone & Henderson, 2009).

1.5.1 Research hypotheses

We initially evaluated *cross-sectional* BFLPEs (see Figure 2a): the compositional effect of Year 1 school-average achievement on students' self-concept in Year 1 (i.e. the BFLPE in Year 1) and the compositional effect of school-average achievement at Year 4 on students' self-concept at the end of Year 4 (i.e. the BFLPE in Year 4). Our hypothesis (*Research Hypothesis 1*) was that the BFLPE would be verified with both Year 1 and Year 4 data (*Research Hypothesis 1a*) and that

adjustments for measurement error would lead to stronger (i.e. more negative) BFLPE estimates (*Research Hypothesis 1b*).

Further, we specified a longitudinal model of compositional effects that looked at the effects of Year 1 school average achievement on students' Year 4 achievement and academic self-concept simultaneously (Becker & Neumann, 2016). The focus was on how corrections for measurement error altered inferences regarding the magnitude and direction of the two effects when these were modelled simultaneously. Our hypothesis (*Research Hypothesis 2; RH2)* was that correcting for the positive bias in BFLPEs and school composition effect estimates would make both effects more negative, if originally negative, or less positive, if originally positive.

1.5.2 Contribution to knowledge

A well-established result in self-concept research is that school-average achievement has negative effects on academic self-concept (the BFLPE; see section 1.1) and that academic self-concept and achievement are positively correlated (reciprocal effects model; REM). Still, school-average achievement is often found to have a positive effect on students' achievements (the peer spillover effect; see section 1.2). Previous studies that acknowledge this contradiction highlight the importance of analysing the consequences of students attending high-achieving schools, considering both academic achievement and ASC as outcomes (Marsh & O'Mara, 2010; Stäbler, et al., 2017; Rindermann & Heller, 2005). Dicke et al. (2018) evaluated the two effects simultaneously, using a large longitudinal sample of US children, and demonstrated that after controlling for measurement error in the underlying data, the BFLPE was negatively affected and the *peer spillover effect* shifted from positive to slightly below zero. Such findings question previous research that did not control for measurement error bias in evaluating the effect of school average achievement on academic achievement and self-concept, and challenge previous policy and school selection decisions based on relevant studies. It is, therefore, critical to establishing whether the results of Dicke et al. (2018) and other, similar studies, are generalisable in different contexts.

The methodological contribution of our study is that it addresses the need to consider measurement error when assessing the effect of school average achievement on students' ASCs and students' achievements. With our longitudinal analysis, we build on studies that have addressed the two effects simultaneously but that failed to control for measurement error bias (Stäbler et al., 2017).

From a substantive point of view, we seek to verify the BFLPE for students as young as six to nine years old. Relatively few studies have looked at the BFLPE with students in the early stages of primary schooling (e.g., Becker & Neumann, 2016; Dicke, et al., 2018; Guo, Marsh, Parker, Dicke & Van Zanden, 2019; Roy, Guay & Valois, 2015; Lohbeck & Mueller, 2017; Marsh et al., 2015; Marsh, Chessor, Craven & Roche, 1995; Tymms, 2001), and especially so for students as young as the first year of primary school.

The theoretical contribution of the study to the analysis of academic self-concept and academic achievement, is that it resolves the apparent contradiction between ASC research findings of the negative effects of school average achievement (BFLPEs) and the educational effectiveness research finding of positive peer spillover effects.

2. Method

2.1 Data Sample

Our data consisted of mathematics achievement and mathematics self-concept measures of 19,059 students from 593 schools. The data, collected from the same students in both Year 1 and Year 4, were longitudinal in nature. They were kindly provided to us by the Performance Indicators at

Primary School (PIPS) project, run by the Curriculum, Evaluation and Monitoring (CEM) centre at Durham University (Tymms et al., 2009). Of the total number of schools involved in the dataset, we based our analysis only on information from schools that participated in both Year 1 and Year 4 educational assessments. We used data on students entering primary school in the academic year 2004 - 2005; this comprised those students who took either their Year 1 assessment in 2005 or their Year 4 assessment in 2008.

2.2 Measures

2.2.1 Mathematics achievement measures. In order to form single scale scores for students' mathematics achievements in Year 1 and Year 4, the average scores were obtained. To formulate multiple indicators for the student-level mathematics achievement measures, item parcelling was used (Little, Cunningham, Shahar & Widaman, 2002). Hence, for Year 1, we created three parcels by averaging every 3rd item; the test originally consisted of 27 items altogether. In the same way, for Year 4, we created four parcels: this test consisted of 36 items in total. In addition to the number of indicators for latent mathematics achievement being significantly reduced, the use of item parcels gave indicators with a distribution better approaching normality, therefore, facilitating normal theory-based estimation (Little, Rhemtulla, Gibson, and Schoemann, 2013; Matsunaga, 2008). Omega estimates of reliabilities of the Year 1(rel = .875) and Year 4 (rel = .921) mathematics achievement measures were relatively high.

2.2.2 Mathematics self-concept measures. The self-concept measures provided by the PIPS tests consisted of five items in Likert-scale form. The items were originally designed to assess the attitudes of the pupils towards mathematics; each had four options from which to choose. The statements could be characterized as hybrids of attitude and self-concept measures but, for the purposes of the present study, all the items were treated as self-concept measures.

Cronbach's alpha estimate of reliability was .614 at Year 1 and .716 at Year 4. In the doubly manifest approach, the scale score for mathematics self-concept is estimated as the average of the self-concept items; in the doubly latent approach, the items themselves are used as multiple indicators. Student-level self-concept measures, as well as mathematics achievement measures, were all standardised in relation to the total sample by subtracting the overall mean and dividing by the overall standard deviation, so that they had a mean of zero and a standard deviation of one.

2.2.3 *Intra-Class Correlation Coefficient (ICC).* The proportion of variance accounted for by the differences between the schools (Intraclass Correlation Coefficient, or ICC) was substantial (Snijders & Bosker, 2012) with both Year 1 (ICC = .179) and Year 4 (ICC = .167) mathematics achievement measures – justifying the use of multilevel modelling (doubly manifest) and multilevel latent variable models (latent manifest, doubly latent; see section 2.5) in our analysis. For Year 1 (ICC = .074) and Year 4 (ICC = .069) mathematics self-concept measures, the ICC was somewhat lower.

2.3 Missing Data

In our analyses, we distinguished between two types of missing data: *unit non-response*, that refers to cases who did not sit the assessment at a particular time point, and, *item non-response*, that refers to cases that took the test but did not respond to some items (Schafer & Graham, 2002).

2.3.1 Missing data with mathematics achievement measures. Students who completed an inadequate number of items in the mathematics achievement section, thereby preventing reliable inferences from being made, were investigated further. We considered the minimum number of items in the mathematics achievement section that a student should have completed before being

included in the analysis as a non-missing case. Any case with five or fewer items attempted in the test was treated as a unit non-response. This resulted in datasets with an even larger number of missing cases than the original files – 2289 (12%) for year one and 1772 (9%) for year four.

2.3.2 Missing data with mathematics self-concept measures. . All students with data on their mathematics achievements had also completed the mathematics self-concept measures. Since most students who participated in either year one or year four assessment completed all the relevant items, there were no serious problems related to item non-response for mathematics self-concept measures. Hence, no cases were treated as unit non-response because of the high rate of item non- response in self-concept data.

2.3.3 The use of multiple imputation to treat missing data. For the treatment of missing data, we used a two-stage Multiple Imputation (MI) procedures to allow for the multilevel structure in our data. Year 1 and Year 4 self-concept measures were included in the same imputation model as mathematics achievement measures. MI involves replacing missing values with a list of two or more simulated values. In this way, plausible alternative versions of the complete data are produced. Each of these is analysed by a complete-data method. Then the results from each imputed dataset are combined to obtain overall estimates and standard errors. The imputation method followed was the fully conditional specification. This is an iterative Markov chain Monte Carlo (MCMC) method that involves a specification of a group of variables to be used in the imputation model –these comprise the variable list – and a specification of several iterations that should be performed before obtaining the imputed values. In each iteration and for each variable in the order specified in the variable list, it fits a univariate model using the variable to be imputed as a dependent variable and all the other variables in the model as predictors. It subsequently imputes missing variables for the variable being fitted. The method repeats the procedure until the specified number of iterations is reached and the imputed values of the final iteration are saved to the imputed dataset.

2.4 Statistical Analysis

We used, in the first instance, a *cross-sectional* compositional analysis model (Figure 2a): we evaluated the BFLPEs separately in the Year 1 and Year 4 data with (latent manifest, doubly latent; Table 1) and without (doubly manifest) corrections for measurement error (*RH1a*; *RH1b*). At a subsequent stage, we specified a *longitudinal model* (Figure 2b), in which we simultaneously modelled the compositional effect of school-average achievement in Year 1 on students' self-concept in Year 4, and on students' mathematics achievement in Year 4, adjusting for both individual achievement and self-concept in Year 1 (*RH2*). All analyses were performed in Mplus 7.4 (Muthén & Muthén, 1998–2015). In comparing estimates of compositional effects across the different approaches, our focus was on the effect size estimate, rather than on the unstandardized estimate; to estimate *effect sizes*, we used the measure recommended by Parker, Marsh, Lüdtke, & Trautwein (2013; Marsh et al., 2009; Nagengast & Marsh, 2012), which is based on the total student-level variance.

[Insert Figure 2a and Figure 2b here]

3. Results

3.1 Cross-Sectional BFLPEs

With respect to assessing the magnitude of the Big-Fish-Little-Pond Effect (BFLPE) for Year 1 (five to six year olds) and Year 4 students (eight to nine year olds), the expectation was that the BFLPE would be verified in both year groups (see *RH1a*); this hypothesis is supported (Table 2). We initially applied the conventional approach to compositional analysis, which does not adjust for measurement error (doubly manifest). A small and marginally significant negative effect of school-average prior achievement was observed with Year 1 data (β_{com}= -.045, se =.022, ES = -.031). A stronger effect was detected using Year 4 data (β_{com} = -.204, se = .025, ES = -.120). Adjustments for measurement error in the individual-level mathematics achievement and the school-level aggregate (the latent manifest approach) resulted in more negative BFLPEs with both year one (β_{com}= -.024, se = .013, ES = -.031) and year four (β_{com}= -.168, se = .022, ES = -.138) data; additional adjustments for sampling error (the double latent model; full-correction approach) led to even more negative estimates of the compositional effect of school-average achievement on students' self-concept in Year 1 and Year 4, with the BFLPEs estimated equal to β_{com}= -.037, (se = .021, ES= -.035) and to β_{com}= -.203, (se = .024, ES = -.154), respectively. Our findings support our research hypothesis (*RH1b*): namely, that adjustments for measurement error would lead to stronger BFLPEs.

3.2 The Effect of School-Average Achievement on Students' Mathematics Achievement and Mathematics Self-concept

The second research hypothesis of our study concerns the impact of measurement error adjustments on inferences regarding the effect of school-average achievement in Year 1 on students' self-concept and achievement in Year 4. To this end, a longitudinal model was specified that allows the estimation of these two effects simultaneously (see Figure 2b). In Table 3, we present the estimates for the structural paths of our model, including only total effect estimates for the compositional effect of the school-average achievement in Year 1 on Year 4 individual achievement and individual self-concept.

3.2.1 The compositional effect of school-average achievement in Year 1 on students' self-concept in Year 4

With the longitudinal multilevel compositional analysis model (doubly manifest approach), the BFLPE was found to be equal to β_{com}= -.182 (se = .026, ES = - 230). After making adjustments for measurement error (latent manifest approach; doubly latent approach), a more negative effect was detected, equal to β_{com}= -.155 (se = .025, ES = -.271) when sampling error is not adjusted for, and equal to β_{com}= -.184 (se = .027, ES = -.292) when it is. Hence, the average achievement of the school's intake in Year 1 – i.e., the achievement composition of the school's intake, continued to influence the students' self-concept several years later, until Year 4.

3.2.2 The compositional effect of school-average achievement in Year 1 on students' achievement in Year 4

A negative and significant – albeit small (β_{com}= -.071, se = .031, ES = -.078) – compositional effect of school-average achievement in Year 1 on students' achievement in Year 4 was detected. As expected, when adjustments for measurement error were made the effect increased in magnitude – that is, negatively– with the latent manifest approach giving an estimate of β_{com}= -.132 (se = .036, ES = -.134), and the doubly latent approach giving an estimate of β_{com}= -.151 (se = .017, ES = -.138). Although the size of these effects is small, the critical issue in relation to the current investigation is that the effects of school-average achievement on subsequent achievement were significantly negative, not positive.

3.3 Additional Analyses: Modelling ASC in Year 4 as a Mediator of the Negative Effect of School Average Achievement on Year 4 Achievement

In a separate analysis, we investigated whether ASC in Year 4 mediated the negative effect of school- average achievement in Year 1 on mathematics achievement in Year 4. From a substantive point of view, this would suggest that a longitudinal BFLPE, manifesting itself in the first four years of primary schooling could, at least in part, explain the occurrence of a negative school compositional effect of Year 1 achievement on students' progress in mathematics from Year 1 to Year 4. Our analyses (Table 4) reveal a negative and significant indirect effect of school-average achievement in Year 1 on students' achievement in Year 4, via students' self-concept measures in the same academic year, both when the doubly manifest ($\beta_{ind} = -.083, se = .032$), and when the latent manifest ($\beta_{ind} = -.029$ $se = .004$) or doubly latent ($\beta_{ind} = -.034, se = .005$) approaches were used. The direct effect of school-average achievement in Year 1 on students' subsequent achievement remained negative and statistically significant; for the doubly manifest approach it was equal to $\beta_{dir} = -.058$ ($se = .032$), while for the latent manifest approach it was $\beta_{dir} = -.117$ ($se = .036$) and, for the doubly latent, $\beta_{dir} = -.137$ ($se = .041$). Therefore, even though the longitudinal BFLPE could explain some of the negative effects of school-average achievement in Year 1 on students' progress in mathematics from Year 1 to Year 4, other factors could also be contributing to the manifestation of this effect (see section 4.3.4).

4. Discussion

4.1 Evaluating BFLPEs in Years 1 to 4: Methodological and Substantive Implications

A focus of our study was to verify BFLPEs for primary years one to four: that is, for students as young as six to nine years of age, and to quantify the impact of measurement error bias in these estimates. We evaluated models that control for measurement error bias (Marsh et al., 2009), and models that do not (Table 1). There is clear evidence for the prevalence of a BFLPE with Year 4 mathematics achievement and self-concept measures (Year 4 cross- sectional BFLPE; Table 2). However, the negative compositional effect of school-average achievement in Year 1 on students' self-concept in Year 1 was relatively small and only marginally significant (Year 1 cross-sectional BFLPE; Table 2). Based on the longitudinal BFLPE model, by Year 4 the BFLPE based on school-average achievement in Year 1 grows much larger – it becomes almost seven times as large – and becomes highly significant (Table 3). Hence, additional BFLPEs occured in the years following primary Year 1 (Marsh, Kong & Hau, 2000). One reason for this could be the fact that social comparison processes and the relation between ASC and achievement are weak in Year 1. Another explanation for the BFLPEs being stronger for students in higher grades could be the fact that students' self- concept becomes more aligned with their achievements as students grow older: this developmental hypothesis could be the focus of future research on relevant topics underlying the BFLPE hypothesis.

We found no evidence of a peer spillover effect (Table 3): the effect of school-average achievement in Year 1 on students' self-concept in Year 4 was small, negative and significant. Our findings replicate those of Televantou et al. (2015) who, using mathematics achievement data on the same sample of students as our study, also revealed a negative compositional effect that became more negative after adjustments for measurement error. However, Televantou et al. failed to evaluate the school compositional effects of average achievement on students' self-concept and based their findings on simpler models.

Given the widespread misconceptions about both the direction and the appropriate methodology for testing peer spillover effects, the basing of our results on a large nationally representative sample of young UK students makes an important contribution to existing research on primary and secondary students in the US (Dicke et al., 2018; Marsh, 1991). Dicke et al., showed that with an appropriate methodology, the biased estimates of the peer spillover effect were transformed from a positive effect (that is consistent with popular belief), to a slightly negative effect. Thus, the main conclusion of the Dicke et al. study was that "the direction of peer spillover effect is not positive, which is most important from a policy perspective" (2018, p. 31). Dicke et al., additionally controlled for pre- existing differences in their models – an issue we did not address in the present study, due to restrictions in our data (see section 4.3.1). We note that, in this respect, the results of our study are much convincing in that, even without controlling for these methodological issues (measurement error and pre-existing differences), the peer spillover effect was negative and the methodological controls only increased this negative effect.

Our – and similar – work can inform the ongoing debates in England that address the impact of segregation (e.g. through parental choice, selective schooling or neighbourhood clustering) on students' educational outcomes (Jenkins, Micklewright & Schnepf, 2008). This is especially so because the prevalence of measurement error in baseline achievement has already been shown to seriously bias English schools' 'progress' value-added measures, used for accountability purposes.

Consequently, the putative superiority of grammar schools' performance must be considered highly ambiguous (Perry, 2018). From a developmental psychology perspective, the prevalence of BFLPEs in primary Year 1 is a remarkable finding, since it implies that the social comparison processes underpinning the BFLPE are evident at such a young age. Obtaining insight into how school composition affects developments in achievement and self-concept in early school years, as proposed in our study, can help shape policies to improve

academic outcomes long term, since
persistent evidence suggests that early schooling can have a big impact on a child's later academic development (Sylva, Melhiush, Sammons, Siraj-Blatchford & Taggart, 2010). For instance, our findings can inform research on the impact of the social context of the school on children's affective and motivational outcomes, e.g. school adjustment (Perry & Weinstein, 1998; Wentzel, Baker & Russell, 2009).

4.2 Positive Assimilation versus Negative Contrast Effects

It is evident that, in our data, the negative frame-of-reference effect is stronger than the positive reflected-glory effect (see section 1.1.1), although the size of each effect is unknown. Our findings are in line with previous research, which has failed to provide evidence for a systematic positive assimilation effect at the school-level (Trautwein, Lüdtke, Marsh, & Nagy, 2008; Trautwein, Lüdtke, Marsh, Köller & Baumert, 2006). For instance, Chmielewski, Dumont & Trautwein, (2013) claim that, while assimilation effects are expected to be stronger in course-by-course tracking and within-school streaming, they are weaker, if existent at all, in between-school streaming. Consistently with our study, Chmielewski et al. (p. 943) report a negative coefficient for school mean mathematics achievement; thus, evidence for contrast effects. It should, however, be clarified that differences in average achievement across schools in our sample are not due to explicit tracking in the Chmielewski study, but rather implicit tracking due to other factors such as social segregation in relation to post code. However, we also note that we know of no BFLPE studies showing that track on its own has a positive effect on academic self-concept. Rather, a few studies have shown that that once the substantial negative effect of school-average ability (the BFLPE) is controlled that there is a small positive effect of track. However, this can be interpreted as merely a positive effect of prior achievement at the individual student level rather than a compositional effect (see discussion by Marsh et al., 2018). Obviously, this is a relevant area for further research (e.g.

Marsh, Parker & Pekrun, 2019) but is not relevant to our study as there was no explicit tracking in primary schools considered here.

4.3 Methodological Limitations

4.3.1 Issues of validity of interpretations in relation to omitted variables

A major focus of the present investigation is the bias that arises in the estimates of compositional effects in assessing the school compositional effects of average achievement and the BFLPE, due to measurement error at level 1. However, at the same time, the insufficiency of level 1 covariates controlled for in compositional models has also been claimed to lead to bias in the estimated compositional effects: the issue of omitted variable bias (Harker & Tymss, 2004). While we deal with measurement error bias, we do not deal with omitted variable bias in our estimation of each of the two compositional effects. The issue of the insufficiency of the control of level 1 covariates is quite distinct from measurement error in the student-level variable on which the aggregate is based and requires a separate approach. Recent work by Dicke et al. (2018) represents one attempt to control for both types of error in the assessment of BFLPEs and peer spillover effects. An alternative approach to omitted variable bias is demonstrated by Caro, Kyriakides and Televantou (2018). In their study, Caro et al. outline an analytical framework to address how omitted prior achievement bias can be corrected for in the context of large-scale assessment when the focus is on the effect of certain teaching strategies on students' achievements. Both studies can be the basis for future research evaluating the magnitude of compositional effects at the level of the school or the classroom, net of measurement error and of omitted variable bias.

When investigating factors associated with between-school differences in their students' outcomes, it is also important to distinguish between compositional effects from institutional effects (Maaz, Trautwein, Lüdtke, & Baumert 2008). The latter refers to the impact of better quality of educational provision typically offered by schools with higher school average achievement (e.g. better trained teachers, resourcing, curricula). Institutional effects typically result in better outcomes for schools with a high school average ability. Our findings, however, suggest this could also be the case if schools with a low school average achievement were given these advantages. Future studies could address relevant issues.

4.3.2 Choosing the correct estimate: The partial- versus the full-correction approach

In a simulation study, Lüdtke et al. (2011) identified conditions under which the convergence and the estimate accuracy of the partial- and full-correction compositional analysis models implied by the Marsh et al. framework may be problematic. However, given the number of students per school in our analysis (on average 32) he number of schools (~500) and the ICC of Year 1 mathematics achievement (ICC~.179), we faced no serious issues with the application of the latent manifest and the doubly latent models, in terms of accuracy or convergence. Lüdtke et al. (2011) advise that, from a statistical perspective, both the latent manifest and the doubly latent approach should be used to obtain a bias-free estimate of the compositional effect; the two estimates should serve as bounds for the true parameter value (Marsh et al., 2009). In a subsequent study, Marsh et al. (2012), distinguish between “contextual” and “climate” level 2 constructs, suggesting that the true value of the compositional effect of school average achievement, a contextual construct, should be closer to the estimate obtained using the latent manifest approach (no adjustments for sampling error); the doubly latent model may over-correct for bias in the respective estimate. Importantly, with respect to our study, the two statistical estimates are not substantially different from each other.

4.3.3 The use of item parcels

For the purposes our study, we made a methodological compromise in that we formed the multiple indicators associated with the set of mathematics achievement tests using item parcels. While item parcelling is a highly debatable practice, its use can be reasonably justified when the focus of a study is on the assessment of the structural paths of latent variable models (Marsh, Lüdtke, Nagengast, Morin & Von Davier, 2013). This is especially the case with our study, given that Year 1 and Year 4 mathematics achievement tests were reasonably unidimensional. By using multiple indicators in the context of the common factor model and of classical test theory, we follow other studies that have used the Marsh et al. (2009) framework (e.g. Nagengast & Marsh, 2012; Televantou et al., 2015). An alternative possibility would be, following Dicke et al. (2018), to go with item response theory, which results in a single score (William & Hazer, 1986); measurement error associated with the derived scores, however, cannot be as easily incorporated into the multilevel latent variable modelling framework.

4.3.4 Limitations of mediation analysis

The final study (see section 3.3) investigated whether the BFLPE (Table 3) – based on school-average achievement in Year 1 – is a mechanism that can explain the occurrence of the negative school compositional effect of school-average achievement on students' mathematics development from Year 1 to Year 4. We tested mediation via self-concept in Year 4 (Table 4): a negative and a statistically significant mediation effect was found, while the direct effect became smaller in size.

Nevertheless, the results of cross-sectional mediation models do not necessarily reflect longitudinal processes (Maxwell, Cole, & Mitchell, 2011): any causal interpretation of the findings is only tentative (Nagengast & Marsh, 2012). In order to fully test the implications of this issue, three or more waves of data would be required (Dicke et al., 2018; Stäbler et al., 2017). Stronger designs, e.g. propensity score matching (Aral, Muchnik & Sundararajan, 2009) or instrumental variables (Aral & Nicolaides, 2017) would be required before claiming causality in the mechanisms identified as potentially underpinning the occurrence of a negative school composition effect.

4.4 Conclusion

Our study verifies the prevalence of a Big-Fish-Little-Pond Effect (BFLPE) in a large sample of English primary Year 1 and Year 4 students, as young as six to nine years of age. We used models that corrected for measurement error bias (doubly latent models); BFLPEs became more negative after adjusting for measurement error. In a longitudinal model that looks simultaneously at the effects of school-average achievement on students' subsequent achievement and ASC, we demonstrated a negative compositional effect for both outcomes. The relative difference in the magnitude of the estimated effects becomes smaller, once measurement error is adjusted. We showed that BFLPEs are one potential mechanism responsible for the occurrence of a negative school compositional effect on school-average achievement, in respect of students' development in mathematics from Year 1 to Year 4. Our findings call into question the supposed advantages of attending higher achievement schools.

References

Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily- driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, *106*(51), 21544- 21549.

Aral, S., & Nicolaides, C. (2017). Exercise contagion in a global social network. *Nature communications*, *8*, 14753.

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

Boonen, T., Speybroeck, S., de Bilde, J., Lamote, C., Van Damme, J., & Onghena, P. (2014). Does it matter who your schoolmates are? An investigation of the association between school composition, school processes and mathematics achievement in the early years of primary education. *British Educational Research Journal*, *40*(3), 441-466. https://doi.org/10.1002/berj.3090

Byrne, B. M. & Shavelson, R.J. (1986). On the structure of adolescent self-concept. *Journal of Adolescent Psychology*, *78*(6), 474-481. http://dx.doi.org/10.1037/0022-0663.78.6.474

Caro, D. H., Kyriakides, L., & Televantou, I. (2018). Addressing omitted prior achievement bias in international assessments: An applied example using PIRLS-NPD matched data. *Assessment in Education: Principles, Policy & Practice*, *25*(1), 5-27. https://doi.org/10.1080/0969594X.2017.1353950

Chmielewski, A. K., Dumont, H., & Trautwein, U. (2013). Tracking effects depend on tracking type: An international comparison of students' mathematics self-concept. *American Educational Research Journal*, *50*(5), 925-957. https://doi.org/10.3102/0002831213489843

Creemers, B. P. M., Kyriakides, L., & Sammons, P. (2010). *Methodological advances in educational effectiveness research*. London: Routledge.

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. http://dx.doi.org/10.1037/edu0000259P

Festinger, L. (1954). A theory of social comparison processes. *Human relations*, *7*(2), 117-140.

Fruehwirth, C.J. (2013). Identifying peer achievement spillovers: Implications for desegregation and the achievement gap. *Quantitative Economics*, 4, 85–124. https://doi.org/10.3982/QE93

Gibbons, S., and Telha, S. (2012). "Peer effects: Evidence from secondary school transition in England." 6455.

Discussion Paper Series, Forschungsinstitut Zur Zukunft Der Arbeit. http://ftp.iza.org/dp6455.pdf

Gray, J., Jesson, D. & Sime, N. (1990). Estimating differences in the examination performances of secondary schools in six LEAs: a multi-level approach to school effectiveness. *Oxford Review of Education, 16*(2), 137-158. http://www.jstor.org/stable/1050400

Guay, F., Larose, S. & Boivin, M. (2004). Academic self-concept and educational attainment level: A ten-year longitudinal study. *Self and Identity, 3*(1), 53-68. https://doi.org/10.1080/13576500342000040

Guo, J., Marsh, H. W., Parker, P. D., Dicke, T., Van Zanden, B. (2019). 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

Harker, R. & Tymms, P. (2004). The effects of student composition on school outcomes. *School Effectiveness and School Improvement, 15*(2), 177-199. http://dx.doi.org/10.1076/sesi.15.2.177.30432

Huguet, P., Dumas, F., Marsh, H., Régner, I., Wheeler, L., Suls, J., ... & Nezlek, J. (2009). Clarifying the role of social comparison in the big-fish–little-pond effect (BFLPE): An integrative study. *Journal of Personality and Social Psychology*, *97*(1), 156. http://dx.doi.org/10.1037/a0015558

Hutchison, D. (1993) School effectiveness studies using administrative data. *Educational Research, 35*(1), 27-47. https://doi.org/10.1080/0013188930350102

Hutchison, D. (2004). The effect of measurement errors on apparent group-level effects in educational progress.

Quality and Quantity, 38(4), 407-424.

https://doi.org/10.1023/B:QUQU.0000043116.22582.fd Hutchison, D. (2007). When is a compositional effect not a compositional effect? *Quality and Quantity, 41*(2), 219-232. https://doi.org/10.1007/s11135-007-9094-2

Jenkins, S. P., Micklewright, J., & Schnepf, S. V. (2008). Social segregation in secondary schools: How does England compare with other countries? *Oxford Review of Education*, *34*(1), 21-37. https://doi.org/10.1080/03054980701542039

Lavy, V. Silva, O. & Weinhardt, F. (2012). The good, the bad, and the average: Evidence on ability peer effects in schools. *Journal of Labor Economics. 30*(2) 367-414. https://doi.org/10.1086/663592

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

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods, 18*, 285-300. https://doi.org/10.1037/a0033266

Lohbeck, A., & Möller, J. (2017). Social and dimensional comparison effects on math and reading self-concepts of elementary school children. *Learning and Individual Differences*, *54*, 73-81. https://doi.org/10.1016/j.lindif.2017.01.013

Lüdtke, O., Marsh, H. W., Robitzsch, A. & Trautwein, U. (2011). A 2× 2 taxonomy of multilevel latent contextual models: Accuracy–bias trade-offs in full and partial error correction models. *Psychological Methods, 16*(4), 444-467.

https://doi.org/10.1037/a0024376

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T. & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods, 13*(3), 203-229. https://doi.org/10.1037/a0012869

Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. Child Development Perspectives, 2(2), 99-106. https://doi.org/10.1111/j.1750- 8606.2008.00048.x

Marks, G. N. (2015). "Are school-SES effects statistical artefacts? Evidence from longitudinal population data." *Oxford Review of Education*, *41*(1). Routledge: 122-144. doi: 10.1080/0305498 5.2015.1006613.

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology,* *79*(3), 280-295. http://files.eric.ed.gov/fulltext/ED278685.pdf

Marsh, H. W. (1991). Failure of high-ability high schools to deliver academic benefits commensurate with their students' ability levels. *American Educational Research Journal*, *28*(2), 445-480. http://www.jstor.org/stable/1162948

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*(1), 258-271. https://doi.org//10.1037/a0037485

Marsh, H. W., Chessor, D., Craven, R. & Roche, L. (1995). The effects of gifted and talented programs on academic self-concept: The big fish strikes again. *American*

Educational Research Journal, 32(2), 285-319. https://doi.org/10.3102/00028312032002285

Marsh, H. W., Chessor, D., Craven, R., & Roche, L. (1995). The effects of gifted and talented programs on academic self-concept: The big fish strikes again. *American Educational Research Journal*, *32*(2), 285-319. https://doi.org/10.3102/00028312032002285,

Marsh, H. W., & Craven, R. G. (2005). A reciprocal effects model of the causal ordering of self-concept and achievement: New support for the benefits of enhancing self-concept. In H. W. Marsh, R. G. Craven, & D. M. McInerny (Eds.), *International advances in self research: The new frontiers of self research* (Vol. 2, pp. 15-51). Greenwich, CT: Information Age.

Marsh, H. W. & Craven, R. G. (2006) Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1*(2), 133-163. https://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W. & O'Mara, A. J. (2010). Long-term total negative effects of school-average ability on diverse educational outcomes. *Zeitschrift für Pädagogische Psychologie, 24*(1), 51-72. https://doi.org/10.1024/1010-0652/a000004

Marsh, H. W. & Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology, 47*(1), 213-231. http://dx.doi.org/10.1037/0022-3514.47.1.213

Marsh, H. W. & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. *Journal of Educational Psychology, 89*(1), 41-54. http://dx.doi.org/10.1037/0022-0663.89.1.41

Marsh, H. W., Kong, C. K. & Hau, K. T. (2000). Longitudinal multilevel models of the big-fish-little-pond effect on academic self-concept: Counterbalancing contrast and reflected-glory effects in Hong Kong schools. *Journal of Personality and Social Psychology, 78*(2), 337-349. https://doi.org/10.1037/0022- 3514.78.2.337

Marsh, H. W., Kuyper, H., Morin, A. J. S., Parker, P. D., & Seaton, M. (2014). Big-fish-little-pond social comparison and local dominance effects: Integrating new statistical models, methodology, design, theory and substantive implications. *Learning and Instruction*, *33*, 50-66. https://doi.org/10.1016/j.learninstruc.2014.04.002

Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J. S., & Von Davier, M. (2013). Why item parcels are (almost) never appropriate: Two wrongs do not make a right – camouflaging misspecification with item parcels in CFA models. *Psychological Methods*, *18*, 257-284. http://www.jstor.org/doi:10.1037/a0032773

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S. & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist,* *47*(2), 106-124. https://doi.org/10.1080/00461520.2012.670488

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B. & Nagengast, B. (2009). Doubly latent models of school contextual effects: Integrating Multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research,* *44*(6), 764- 802. https://doi.org/10.1080/00273170903333665

Marsh, H. W., Parker, P. D., Guo, J., Pekrun, R., & Basarkod, G., (2020). Psychological comparison processes and self-concept in relation to five distinct frame-of-reference effects: Pan-human cross-cultural generalizability over 68 countries. *European Journal of Personality.* Online version https://doi.org/10.1002/per.2232

Marsh, H. W., Pekrun, R., Murayama, K., Arens, A. K., Parker, P. D., Guo, J. & Dicke, T. (2018). An integrated model of academic self-concept development: Academic self-concept, grades, test scores, and tracking over 6 years. *Developmental Psychology*, *54*(2), 263-280. https://doi.org/10.1037/dev0000393

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*(3), 231-242. https://doi.org/10.1027/1016-9040/a000332

Marsh, H. W., Seaton, M., Kuyper, H., Dumas, F., Huguet, P., Régner, I., Buunk, 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

Matsunaga, M. (2008). Item parceling in structural equation modeling: A primer. *Communication Methods and Measures,* *2*(4), 260-293. https://doi.org/10.1080/19312450802458935

Maxwell, S. E., Cole, D. A., & Mitchell, M. A. (2011). Bias in cross-sectional analyses of longitudinal mediation: Partial and complete mediation under an autoregressive model. *Multivariate Behavioral Research*, *46*(5), 816-841. https://doi.org/10.1080/00273171.2011.606716

Muthén, L. K., & Muthén, B. O. (1998–2015). *Mplus user's guide* (7th ed.). Los Angeles, CA: Authors.

Nagengast, B., & Marsh, H. W. (2012). Big fish in little ponds aspire more: Mediation and cross-cultural generalizability of school-average ability effects on self-concept and career aspirations in science. *Journal of Educational Psychology*, *104*(4), 1033. https://doi.org/10.1037/a0027697

Nash, R. (2003). Is the school composition effect real? A discussion with evidence from the UK PISA data. *School effectiveness and school improvement*, *14*(4), 441-457. https://doi.org/10.1076/sesi.14.4.441.17153

Parker, P. D., Marsh, H. W., Lüdtke, O., & Trautwein, U. (2013). Differential school contextual effects for math and English: Integrating the big-fish-little-pond effect and the internal/external frame of reference. *Learning and Instruction*, *23*, 78-89. https://doi.org/10.1016/j.learninstruc.2012.07.001

Pekrun, R., Murayama, K., Marsh, H. W., Goetz, T., & Frenrzel, A. C. (2019). Happy fish in little ponds: Testing a reference group model of achievement and emotion. *Journal of Personality and Social Psychology.,* *117*(1), 166-185. https://doi.org/10.1037/pspp0000230

Perry, K. E., & Weinstein, R. S. (1998). The social context of early schooling and children's school adjustment. *Educational Psychologist*, *33*(4), 177-194. https://doi.org/10.1207/s15326985ep3304_3

Perry, T. (2018). Phantom compositional effects in English school value-added measures: The consequences of random baseline measurement error. *Research Papers in Education*, 1-24. https://doi.org/10.1080/02671522.2018.1424926

Pokropek, A. (2015). Phantom effects in multilevel compositional analysis: Problems and solutions. *Sociological Methods & Research*, *44*(4), 677-705. https://doi.org/10.1177/0049124114553801

Rindermann, H., & Heller, K. A. (2005). The benefit of gifted classes and talent schools for developing students' competences and enhancing academic self-concept. *Zeitschrift für Pädagogische Psychologie*, *19*(3), 133-136. https://doi.org/10.1024/1010-0652.19.3.133

Roy, A., Guay, F., & Valois, P. (2015). The big-fish-little-pond effect on academic self-concept: the moderating role of differentiated instruction.

https://doi.org/10.1016/j.lindif.2015.07.009

Schafer, J.L. & Graham, J.W. (2002) Missing data: our view of the state of the art. *Psychological methods,* 7 (2), 147-177. https://doi.org/10.1037/1082-989X.7.2.147

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

Smith, D., & Tomlinson, S. (1989). *The school effect.* Lancaster: Policy Studies Institute, University of Lancaster.

Snijders, T. A. B. & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modelling (*2nd ed.). London: SAGE.

Stäbler, F., Dumont, H., Becker, M., & Baumert, J. (2017). What happens to the fish's achievement in a little pond? A simultaneous analysis of class-average achievement effects on achievement and academic self- concept. *Journal of Educational Psychology*. *109*(2), 191-207. http://dx.doi.org/10.1037/edu0000135

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (Eds.). (2010). *Early childhood matters: Evidence from the Effective Pre-school and Primary Education Project*. London: Routledge.

Teddlie, C., Stringfield, S. & Reynolds, D. (1999). Context issues within school effectiveness research. In C. Teddlie and D. Reynolds (Eds.), *The international handbook of school effectiveness research* (pp. 160- 187). London: Falmer Press.

Televantou, I., Marsh, H. W., Kyriakides, L., Nagengast, B., Fletcher, J. & Malmberg, L-E. (2015). Phantom effects in school composition research: consequences of failure to control biases due to measurement error in traditional multilevel models. *School Effectiveness and School Improvement.* *26*(1), 75-101.

https://doi.org/10.1080/09243453.2013.871302

Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O. & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology, 98*(4), 788-806. https://doi.org/10.1037/0022- 0663.98.4.788

Trautwein, U., Lüdtke, O., Marsh, H. W., & Nagy, G. (2008). *Within-school social comparison: Perceived class prestige predicts academic self-concept.* Berlin: Max Planck Institute for Human Development.

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*(2), 161-181. https://doi.org/10.1076/sesi.12.2.161.3452

Tymms, P., Jones, P., Albone, S., & Henderson, B. (2009). The first seven years at school. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 67-80. https://doi.org/10.1007/s11092-008-9066-7

Wentzel, K., Baker, S., & Russell, S. (2009). Peer relationships and positive adjustment at school. In M. J. Furlong,

R. Gilman, & E. S. Huebner (Eds.), *Handbook of positive psychology in schools* (pp. 229-243). London: Routledge.

Williams, L. J., & Hazer, J. T. (1986). Antecedents and consequences of satisfaction and commitment in turnover models: A reanalysis using latent variable structural equation methods. *Journal of applied psychology*, *71*(2), 219-231.

Willms, J. D. (1985). The balance thesis: contextual effects of ability on pupils' O-grade examination results.

Oxford Review of Education, *11*(1), 33-41. https://doi.org/10.1080/0305498850110103

Woodhouse, G., Yang, M., Goldstein, H. & Rasbash, J. (1996) Adjusting for measurement error in multilevel analysis. *Journal of the Royal Statistical Society. Series A (Statistics in Society), 159*(2), 201-212. https://doi.org/10.2307/2983168

Zirkel, P. A. (1971). Self-concept and the "disadvantage" of ethnic group membership and mixture. *Review of Educational Research*, *41*, 211-225. http://www.jstor.org/stable/1169528

1. Tables

Table 1. *The use of Marsh et al.'s (2009) multilevel latent modelling framework to correct for measurement error bias*[1]

<table>
<tr><td rowspan="5">Measurement Error Adjustments</td><td>No</td><td colspan="2">Doubly Manifest Approach:
• Single manifest indicators (one per factor)
• Manifest aggregation of L1constructs to form L2 constructs</td></tr>
<tr><td rowspan="4">Yes</td><td colspan="2">Sampling Error Adjustments</td></tr>
<tr><td>No</td><td>Yes</td></tr>
<tr><td>Latent Manifest Approach:</td><td>Doubly Latent Approach:</td></tr>
<tr><td>• Multiple Indicators (constructs are latent in relation to items)
• Manifest aggregation of L1 indicators to form L2 indicators</td><td>• Multiple Indicators (constructs are latent in relation to items)
• Latent aggregation of multiple L1 indicators to form multiple L2 indicators</td></tr>
</table>

notes. [1]Adapted from "Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error", by Marsh et al., 2009. [2]In this table, L1 is used to denote student-level (level 1) variables and L2 is used to denote school- level (level 2) variables.

Table 2. *The effect of school average achievement on students' mathematics self-concept: Cross-sectional models for the Big-Fish-Little-Pond-Effect in Year 1 and Year 4*

	Year 1						Year 4					
	□ *within*		□*com*		Residual variance		□ *within*		□*com*		Residual variance	
					Level 1	Level 2					Level 1	Level 2
	Estimate	*SE*	Estimate	SE			Estimate	SE	Estimate *ES*□*cont*	SE		
Doubly Manifest	.084***	.006	- .045* *ES*□*cont* = - .031	.022	.335	.028	.163	.007	- .204*** *ES*□*cont* = - .120	.025	.415	.031
Latent Manifest	.057***	.005	-.024· *ES*□*cont* = -.031·	.013	.126	.012	.134***	.006	- .168*** *ES*□*cont* = - .138	.022	.199	.019
Doubly Latent	.76***	.006	-.037· *ES*□*cont* = -.035·		.126	.012	.135	.006	- .203*** *ES*□*cont* = - .154	.024	.199	.018

· p <.1, *p < .05, ** p <.01, *** p < .001

Note. □*within* denotes the effect of the individual-level predictor (mathematics achievement) on self-concept; □*cont* denotes the effect of school average achievement on mathematics self-concept (i.e. the big-fish-little-pond-effect estimate); *SE* is the standard error of the parameter estimate; the effect size estimate for the corresponding effect is denoted by *ES*□*cont* .

Table 3. *The Big-Fish-Little-Pond-Effect and the school composition effect: longitudinal analysis*

	The Big Fish Little Pond Effect						***The School Composition Effect***					
	Year 4 Academic Self-Concept on Year 1 Academic Achievement						Year 4 Academic achievement on Year 1 Academic Achievement					
	Doubly Mani		Latent Manifest		Doubly Latent		Doubly Manifest		Latent Manifest		Doubly Latent	
	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*
Level 1: Individual-level predictors												
ACH Y1	.087***	.007	.07***	.006	.07***	.006	.684***	.007	.846***	.011	.846***	.011
ASC Y1	.180**	.009	.259***	.018	.259***	.018	.022*	.008	.024	.018	.023***	.018
Level 2: School-level predictors												
ACH Y1[1]	-.182***	.026	-.155***	.025	-.184***	.027	-.071*	.031	-.132***	.036	-.151***	.017
	ES = **-.230**		*ES* = **-.271**		*ES* = **-.292**		*ES* = **-.078***		*ES* = **-.134**		*ES* = **-.138**	
Residual Variance												
Level 1	.414		.195		.195		.318		.213		.212	
Level 2	.03		.017		.017		.072		.07		.072	

Note. In this table we report the total effect of school average achievement in Year 1 on the respective outcome. ES_{cont} is the effect size estimate of the compositional effect; SE is the standard error

Table 4. *The compositional effect of school average achievement in Year 1 on students' mathematics achievement in Year 4: The indirect effect via mathematics self-concept in Year*

	Doubly Manifest		Doubly Latent	
	Estimate	*SE*	Estimate	*SE*
Direct effect	-.058·	.032	-.137***	.041
Indirect effect	-.025***	,004	-.034***	.005
Total Effect	-.083**	.032	-.169***	.041
	ES□*ttot* = -.09	,034	*ES*□*ttot* = -.157	.038

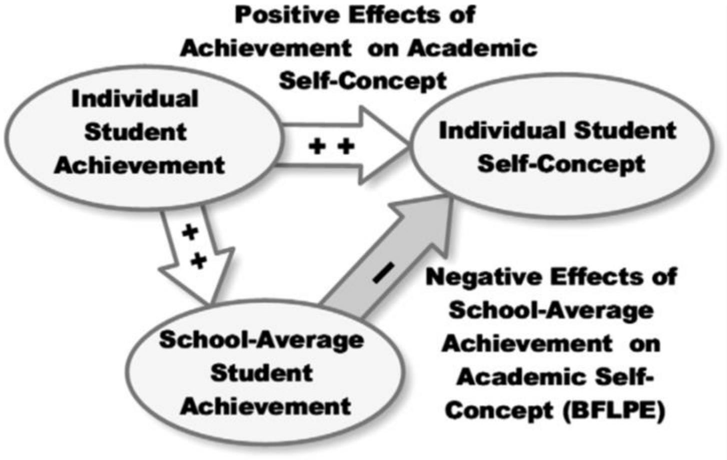
4

·p<.1, *p<.05, **p<.05, ***p<.001

Note. *ES*□*ttot* is the effect size; *SE* is the standard error of the estimate.

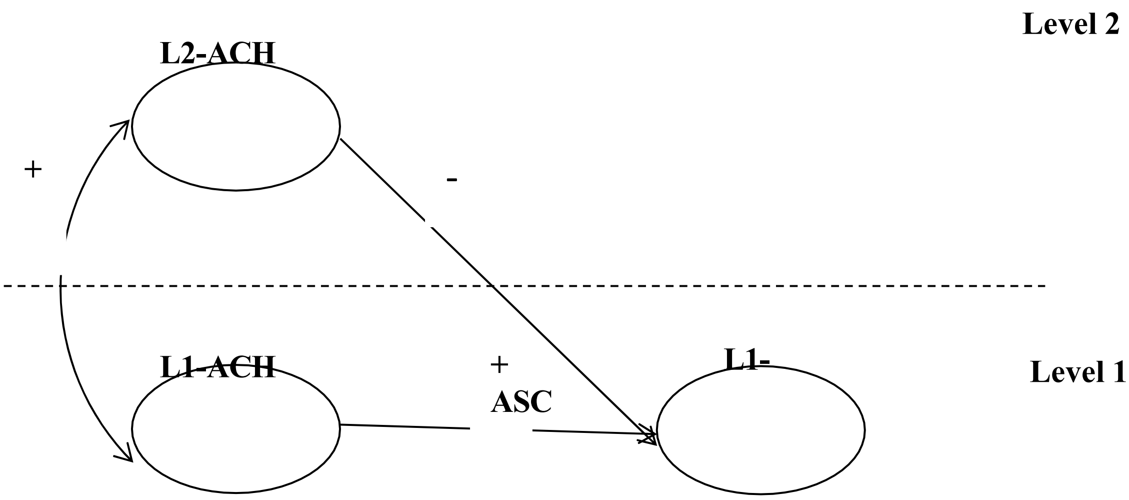
Early BFLPEs on self-concept and academic achievement

Figure 1. The Big-Fish-Little-Pond-Effect (from Marsh, 2007).



Note. The Big-Fish-Little-Pond-Effect: Individual student achievement has a positive effect (++) on academic self-concept, while the effect of school or class average achievement is negative (-).

Figure 2a. Theoretical structural model of the compositional effect of school average achievement on individual self-concept: The cross-sectional Big-Fish-Little-Pond-Effect

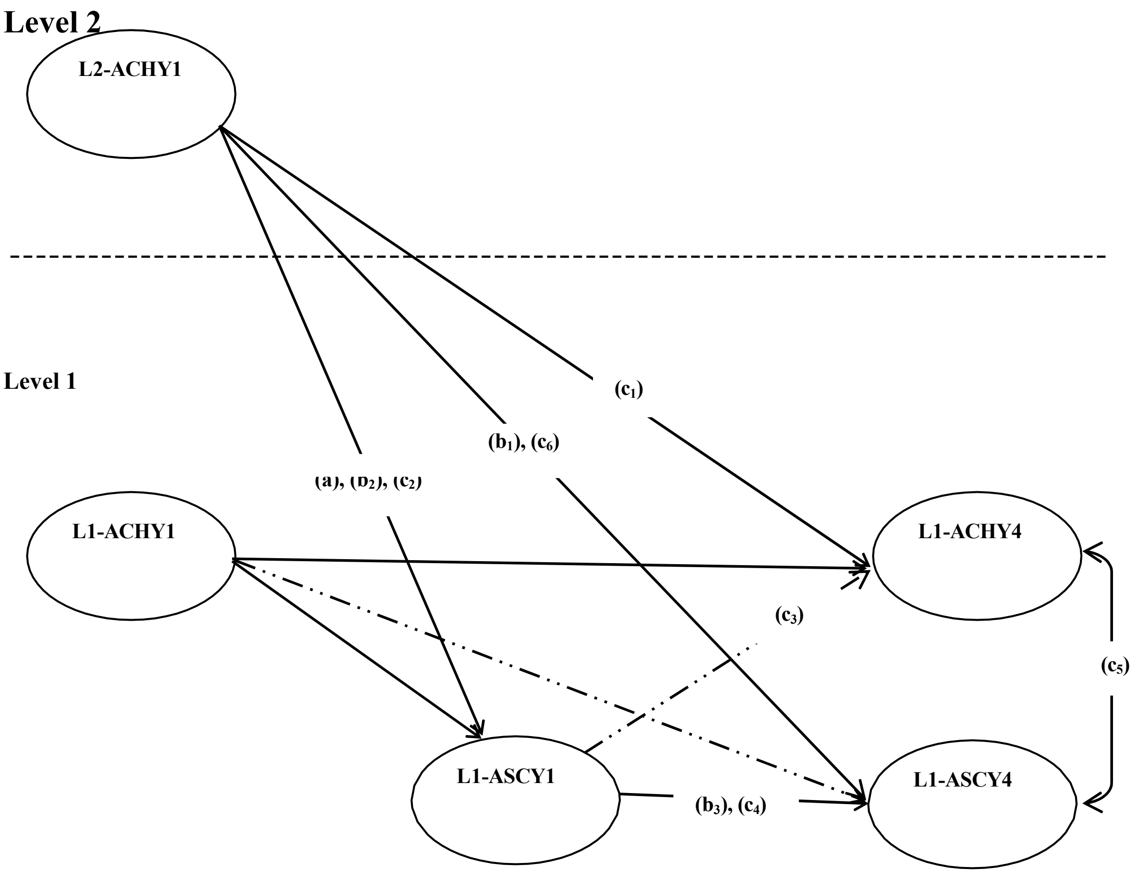


Note. The abbreviation L2-ACH is used to denote the aggregated achievement (school-level average achievement) while L1-ACH is used to denote student-level achievement and L1-ASCs to denote student-level self-concept. All variables are measured *at the same time point.* Plus signs indicate a positive effect; minus signs denote a negative effect.

31

Figure 2b. A longitudinal model describing the Big-Fish-Little-Pond-Effect (BFLPE) in Year 1, the BFLPE in Year 4 and the compositional effects of Year 1 achievement on individual achievement in Year 4.

Year 1 **Year 4**



Note. Model (a) is the BFLPE in Year 1, and involves estimation of the effect of school average achievement in Year 1 (L2-ACH1) on self-concept in Year 1 (L1-ASCY1) after adjustments for achievement in Year 1 (L1-ACH1); only a direct effect (a) is involved in this model. Model (b), the longitudinal BFLPE in Year 4, involves

estimation of both the direct effect (b1) of L2-ACH1 on self-concept in Year 4 (L1-ASCY4), and the indirect effect, through L1-ASCY1 (b2*b3). Model (c), the longitudinal model for the school composition effect, involves estimation of the following effects: the direct effect of L1-ACHY1 on student achievement in Year 4 (L1-ACHY4; c1), the indirect effect of L1-ACHY1 on L1-ACHY4 via L1-ASCY1 (c2*c3) and the indirect effect of L1-ACHY1 on L1-ACHY4 via L1-ASCY4 (c5*c6* c7
+ c5*c6). In order to estimate the latter, we assume a one-directional arrow from L1-ASCY4 to L1-ACHY4