The Formation of Academic Self-concept in Elementary Education:

A Unifying Model for External and Internal Comparisons

Pinxten Maarten¹, Wouters Sofie² (joint first author), Preckel Franzis³, Niepel Christoph⁴, De
Fraine Bieke¹, & Verschueren Karine²

¹ Faculty of Psychology and Educational Sciences, Educational Effectiveness and Evaluation,
KU Leuven, Belgium

² Faculty of Psychology and Educational Sciences, School Psychology and Child and
Adolescent Development, KU Leuven, Belgium

³ Department of Psychology, University of Trier, Germany

⁴ Education, Culture, Cognition and Society (ECCS), Faculty of Language and Literature,
Humanities, Arts and Education, University of Luxembourg, Luxembourg

Sofie Wouters is a postdoctoral fellow of the Research Foundation – Flanders (FWO).

Correspondence concerning this article should be addressed to Sofie Wouters. E-mail:

sofie.wouters@ppw.kuleuven.be

N.B.: When citing this work, cite the original article:

Pinxten, M., Wouters, S. (joint first author), Preckel, F., Niepel, C., De Fraine, B., &
Verschueren, K. (2015). The formation of academic self-concept in elementary education:
A unifying model for external and internal comparisons. *Contemporary Educational
Psychology*. doi: <http://dx.doi.org/10.1016/j.cedpsych.2014.12.003>

Abstract

Given its eminent role in student learning and development, it is important to understand how academic self-concept (i.e., how one perceives oneself in an academic context) is formed. Both internal and external comparisons are considered crucial antecedents: Students form their academic self-concept to a considerable extent by (externally) comparing themselves with others and by (internally) comparing their own performance in different academic domains. Building on previous research in secondary education, the main goal of this study is to test a model integrating both comparison processes in elementary education using a large sample of Grade 4 students ($N = 4,436$) nested in 241 classes. Including the proposed internal and external reference effects in one integrated model, the study provided evidence for the presence of both comparison effects on two academic self-concept domains (i.e., math and verbal self-concept). Specifically, students' achievement in one domain was positively related to self-concept in that domain and negatively related to self-concept in the other domain. Additionally, class-average achievement was negatively related to academic self-concept within each domain and positively across domains. In conclusion, this study stresses the need for further integration of the major models on academic self-concept formation in a unifying theoretical framework.

Key-words: academic self-concept; internal comparison; external comparison; elementary education; internal-external frame of reference model; big-fish-little-pond effect.

1. Introduction

Academic self-concepts (ASCs) comprise mental representations of one's abilities in academic domains. Current models of ASC (e.g., Marsh/Shavelson Model of Marsh, 1990a; nested Marsh/Shavelson Model of Brunner et al., 2010) differentiate between general ASC and domain-specific ASCs. While general ASC reflects an individual's evaluation of his or her academic abilities across subjects ("I am good at most school subjects"), domain-specific ASCs reflect an individual's impression of his or her ability in a specific academic domain, such as mathematics ("I am good at mathematics") or English ("I am good at English"). In the present study, we will focus on the domain- or course-specific level of ASC. ASC tends to be highly correlated with academic self-efficacy as one's confidence in successfully carrying out a specific academic task (Bong & Skaalvik, 2003; Ferla, Valcke, & Cai, 2009). However, different from self-efficacy, ASC is more likely to be rooted in social comparisons. In judging one's ASC, individuals use the achievements of relevant others (in school, usually their classmates) as a frame of reference to evaluate their own achievement level (Seaton, Marsh, & Craven, 2010). When evaluating one's efficacy at a particular task, however, individuals assess their chance of succeeding relative to the task at hand which reduces the influence of frame of reference effects (Seaton et al., 2010).

ASC shows beneficial effects on a wide range of educational outcomes. It positively predicts not only student achievement (e.g., Guay, Marsh, & Boivin, 2003; Marsh & O'Mara, 2008; Pinxten, Marsh, De Fraine, Van Den Noortgate, & Van Damme, 2014), but also academic adjustment (e.g., Wouters, Germeijs, Colpin, & Verschueren, 2011), coursework selection (e.g., Dickhäuser, Reuter, & Hilling, 2005), or interest (Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005). Given its eminent role in student learning and development, it is important to understand how ASC is formed.

Many studies have demonstrated that ASC and academic achievement are positively and reciprocally related to each other (e.g., Huang, 2011; Marsh & Martin, 2011): ASC and achievement are mutually reinforcing in their development so that for equally achieving students the one with the higher ASC is more likely to develop higher achievement over time. However, not only one's actual performance is considered to be of importance. Research has shown that comparisons are crucial antecedents of ASC (Huguet, et al., 2009; Marsh & Martin, 2011; Möller & Marsh, 2013). First, students conduct social comparisons and compare themselves with others (i.e., *external comparisons*). Depending on the achievement level of the respective comparison or reference group, the ASC of equally able students will vary (i.e., ASC decreases with an increasing achievement level of the reference group). Second, students compare their own achievements in different academic domains (i.e., *internal comparisons*). That is, high achievement in one academic domain (e.g., mathematics) can exert a negative influence on the ASC in another academic domain (e.g., English) (for a meta-analysis see Möller, Pohlmann, Köller, & Marsh, 2009).

Although empirical research has unequivocally proven the relevance of both external and internal comparisons for academic self-concept formation, very few studies have empirically considered their combined impact (see Chiu, 2012 and Parker, Marsh, Lüdtke, & Trautwein, 2013 for two recent exceptions). Moreover, none of these studies investigated elementary school students. Yet, it has been suggested that external comparisons may be less important for younger children and that domain-specific self-concepts of younger children may be more strongly interrelated across different academic domains (Dijkstra, Kuyper, van der Werf, Buunk, & van der Zee, 2008; Möller et al., 2009). The present study aims to extend research by studying the complex interplay between internal and external comparisons on students' math and verbal self-concept in a large sample of elementary school students. The impact of internal comparisons on ASC is usually investigated in the internal/external frame

of reference model (I/E Model; Marsh, 1986), the impact of external comparisons or reference groups effects on ASC in the big-fish-little-pond model (BFLP Model; Marsh, 1984). We will first describe each of these models in detail.

2. The Internal/External Frame of Reference Model (I/E Model)

The I/E Model (see Figure 1a), first proposed by Marsh (1986), primarily describes internal, ipsative-like processes wherein students compare their own achievement in one academic domain with their own achievement in other academic domains (i.e., identifying their strengths and weaknesses across different academic domains). The common depiction of the I/E Model is presented in Figure 1a. What can be deduced from Figure 1a is that the I/E model predicts: (a) positive within-domain effects¹ (e.g., a higher math achievement results in a higher math self-concept), and (b) negative cross-domain effects (e.g., a higher math achievement leads to a lower verbal self-concept).

With regard to the correlations between the variables in the model, Marsh (1986) initially hypothesized that the substantial positive correlation between math and verbal abilities would also lead to a positive correlation between math and verbal self-concept. As different academic subjects are often highly interrelated, self-concepts across academic domains should be positively related (Möller et al., 2009). However, because internal comparisons are ipsative, an increase in self-concept in one domain should trigger a decrease in self-concept in another domain, which would entail a negative correlation between self-concepts across domains (Parker et al., 2013). When combined, both processes then result in the observed close to zero correlations between self-concepts in different domains (Marsh, 1986).

A plethora of studies support the predictions of the I/E Model as depicted in Figure 1a and the cross-cultural generalizability of the model (Marsh & Hau, 2004). Möller and colleagues (2009) point at the diverse and sound empirical evidence supporting this model

(i.e., experimental studies, longitudinal studies, introspective diary studies, and meta-analyses). Still, we noticed a few important limitations in research on the I/E model. First, as the name of the model suggests, the I/E Model also mentions external comparison processes as influencing students' ASC (Marsh, 1986). However, both in its original depiction (Marsh, 1986, p. 134) and in later elaborations over the years (e.g., Marsh & Hau, 2004; Skaalvik & Rankin, 1995; Skaalvik & Skaalvik, 2002), the social comparison or external component of the model was only implicitly assumed, but not actually modeled: the achievement level of classmates or schoolmates, for instance, was never included as an actual predictor. Second, studies with elementary school children are scarce. In the meta-analysis of Möller and colleagues (2009), only 10 of the 69 studies under consideration had participants with less than 6 years of schooling. Although the predictions of the I/E Model seemed to generalize over age groups, (positive) correlations between math and verbal self-concept were somewhat larger for younger (before Grade 7) than for older students (Möller et al., 2009). This might indicate weaker effects of internal comparisons (or weaker negative cross-domain effects) for elementary school students than for older students.

3. The Big-Fish-Little-Pond Model (BFLP Model)

In contrast to the I/E Model, the BFLP Model (see Figure 1b for the verbal domain and Figure 1c for the math domain) explicitly focuses on social or external comparisons. Developed by Marsh (1984), it assumes that the comparison of one's academic performance with that of one's immediate peers is an important determinant of ASC and is thus heavily based on social comparison theory (Festinger, 1954). Accordingly, objectively equal accomplishments may be perceived as disappointing by some students leading to a negative ASC when comparing with higher achieving peers, but as successful by others leading to a positive ASC when comparing with lower achieving peers.

Throughout the years, many empirical studies have confirmed the central hypothesis of the BFLP model that class- or school-average achievement is negatively related to students' ASC (when controlling for the positive effect of individual achievement on ASC) (e.g., Köller, Zeinz, & Trautwein, 2008; Marsh & Hau, 2003; Marsh, Köller, & Baumert, 2001; Marsh, Kong, & Hau, 2000; Marsh, Trautwein, Lüdtke, Baumert, & Köller, 2007; Preckel & Brüll, 2010; Preckel, Zeidner, Goetz, & Schleyer, 2008; Thijs, Verkuyten, & Helmond, 2010; Seaton et al., 2010; Wouters, Colpin, Van Damme, De Laet, & Verschueren, 2013). This negative effect has been found to be particularly robust across a wide range of individual and contextual characteristics, educational settings, and countries (Marsh et al., 2008). Nevertheless, there are some pertinent gaps in BFLP research. (1) The large majority of BFLP studies tested the model premises within a single academic domain. As such, internal comparison processes across different academic domains (which are central in the I/E Model) were largely ignored. (2) With a few exceptions (e.g., Köller et al., 2008; Thijs et al., 2010; Wouters et al., 2013), BFLP studies have mainly focused on samples of secondary school students. It is widely accepted that with increasing age students become more susceptible to social comparison information as a means of self-evaluation (Dijkstra et al., 2008). Thus, more research on how the BFLP processes operate at the elementary school level is needed.

4. The Revisited I/E Frame of Reference Model

Despite the term I/E model and although Marsh (1986) originally stipulated the crucial interplay between external (between-schools or classes) and internal (between-domains) comparison processes in academic self-concept formation, the 'external' part of the I/E model has been largely ignored in empirical tests without any further questioning. Interestingly, the BFLP Model outlined above does empirically focus on this external component. Nevertheless, the BFLP Model and the I/E Model have been used and tested completely parallel to each other for many years. Therefore, we aim to integrate both models in one unifying model.

Rather than labeling such a unifying model as, for example, the I/E-BFLP Model, we choose to introduce it as the *Revisited Internal/External Frame of Reference Model (Revisited I/E Model)* (see Figure 1d).

To date, only few studies explicitly combined the I/E and the BFLP models (e.g., Chiu, 2012; Marsh, 1990b, 1994; Parker et al., 2013). In his early studies, Marsh (1990b, 1994) included the effect of school-average achievement (i.e., the effects of external comparisons) into the I/E Model and found support for the predictions of both models – the I/E and the BFLP model. Thus, individual student achievement positively predicted self-concept in matching domains, individual student achievement negatively predicted self-concept in non-matching domains, and school-average achievement negatively predicted students' self-concept in matching domains. Although not predicted a priori, in both studies small positive cross-domain effects of school-average achievement on ASC were observed, meaning that school-average math achievement had a weak positive effect on English self-concept and school-average English achievement had a weak positive effect on math self-concept. According to Marsh, these positive counterbalancing effects may mitigate some of the negative within-domain effects of the BFLP effect. However, these early studies examined these relations only at the student level. To test the central predictions of the BFLP Model, multilevel models that enable researchers to simultaneously test relationships at different levels of the data are needed (Parker et al., 2013). For example, when testing the BFLP effect, researchers are interested in how class- or school-average achievement relates to individual students' ASC after having controlled for the relation of individual student achievement on individual students' ASC. That is, in Marsh' early studies, higher level predictors (i.e., the school-average achievement variables) were not included at the appropriate level (i.e., the school level). Therefore, results of these early studies should be interpreted with caution.

More recently, two studies empirically juxtaposed the I/E with the BFLP Model using more advanced statistical approaches (Chiu, 2012; Parker et al., 2013). Using the TIMSS 2003 database, Chiu (2012) investigated the cross-national generalizability of the unifying model in the math and science domain in a large sample of Grade 8 students. In order to account for the complex structure of the data, the author used a three-level (students, schools, countries) hierarchical regression model. Chiu found strong support for the predictions of the unifying model across 27 countries. However, Chiu did not use a latent modeling approach, typically used when examining latent constructs such as ASC. Additionally, due to multicollinearity problems resulting from the extremely high correlation between school-average mathematics achievement and school-average science achievement ($r = .96$), school-average achievement in both domains was not included in a single model. As a result, Chiu was unable to corroborate Marsh's (1990b, 1994) findings regarding the positive compensatory paths (i.e., positive cross-domain effects of school-average achievement on ASC). In order to overcome this multicollinearity problem, Chiu recommended carefully selected sampling procedures in future models, for example, by only including participants with differential achievement profiles between two different domains (e.g., schools with gifted students in a particular domain versus schools with mixed ability students).

Parker et al. (2013) explicitly addressed this recommendation of Chiu by comparing the results of the unifying model across two school types in the German school context. More specifically, the authors examined a large sample of final year students nested in 157 university track schools. In Germany, one distinguishes two types of schools within this university track: general (or traditional) and themed (or magnet) schools. Themed schools offer a specialized set of programs in one academic domain, preparing students for a career in this particular field/domain. Students are primarily selected based on a narrowly defined area of excellence (e.g., math *or* literature). In contrast, general schools are more selective in a way

that students are selected on the basis of their general academic excellence (e.g., math *and* literature). As such, school-average achievement in the math and verbal domain were more differentiated in themed rather than in general schools (Parker et al., 2013). More specifically, school-level correlations of .74 and .39 between both domains were observed in general and themed schools, respectively.

Regarding the traditional I/E and BFLP Model separately, a similar pattern of results was found for traditional and themed schools. However, when both models were combined in a unifying model, interesting differences emerged. For both school types, the authors found strong negative effects of school-average achievement on students' ASC in the same domain (i.e., BFLP effects). However, cross-domain effects of school-average ability on ASC differed slightly. For themed schools, a positive compensatory effect of school-average math achievement on verbal self-concept and of school-average English achievement on math self-concept was supported. In traditional schools, these relations were not significant.

5. The Present Study

To date, there is some empirical evidence that both the I/E model and the BFLP Model are generally confirmed in elementary school student samples. Nevertheless, research focusing on testing these models in younger students is still scarce. Yet, it has been argued that external comparisons may be less important for younger children as they increase throughout elementary school (Dijkstra et al., 2008). In addition, ASCs of younger children seem to be more strongly interrelated across different academic domains (Möller et al., 2009).

Additionally, cross-cultural findings showed that the strongest BFLP and I/E effects are observed in the German educational context (Marsh & Hau, 2004; Parker et al., 2013), suggesting that more research in other educational settings is needed in order to corroborate German research findings. Therefore, the present study aims at filling both gaps by empirically testing the central predictions of the Revisited I/E Model on a large sample of

Grade 4 students in Flanders, the Northern Dutch-speaking region of Belgium. As the classroom seems the most relevant reference frame in (Flemish) elementary education, we particularly focused on external comparisons with classmates.

Careful synthesis of existing literature led us to the formulation of the following research hypotheses:

1. We expected to find support for the I/E Model in Grade 4: Academic achievement and ASC are positively related *within* a domain (positive within-domain effects, i.e., math achievement is positively related to math self-concept and verbal achievement is positively related to verbal self-concept), but academic achievement and ASC are also negatively related *across* domains (negative cross-domain effects, i.e., math achievement is negatively related to verbal self-concept and verbal achievement is negatively related to math self-concept)
2. We expected to find support for the BFLP effect in Grade 4: Individual math achievement has a positive effect on math self-concept and, when controlling for this effect, class-average math achievement has a negative effect on math self-concept. The same pattern is expected for the verbal domain.
3. We expected that the central predictions of both the I/E Model and the BFLP Model would hold when both models are combined in the unifying model. The existence of positive cross-domain compensatory effects of class-average math achievement on verbal self-concept and of class-average verbal achievement on math self-concept was explored in this unifying model because the research base is too weak to justify a priori hypotheses.
4. We also had some additional hypotheses on the correlations between our main variables. In line with previous research (e.g., Parker et al., 2013), we expected significant positive correlations between verbal and math achievement at both the

individual and the group level. Additionally, based on previous research (see Möller et al., 2009), we expected no significant or only a small positive correlation between verbal and math self-concept.

6. Data and Methods

6.1. Participants

6.1.1. Sample. Data analyzed in the present study stem from the SiBO project (Dutch acronym for School Trajectories in Primary Education; Maes, Ghesquière, Onghena, & Van Damme, 2002). One of the project's main goals was to gain insight into the educational trajectories of students throughout the course of elementary education. In the present study, data were analyzed when students were in Grade 4 (4,436 students in 241 classes; on average 18.4 students per class)². In this large-scale quasi-representative sample, boys and girls were equally distributed (50% - 50%). In Flanders, classes in elementary education are heterogeneously grouped which means that no tracking/streaming practices on the basis of differential abilities are applied. Students typically take lessons in classes, taught by a single teacher for all subjects who stays with the same students throughout the school year.

6.1.2. Data collection procedure. All tests and questionnaires were collectively administered in the classroom at the end of each school year (May – June) by means of paper and pencil method. Questionnaires assessing students' math and verbal self-concept were administered in conjunction with the achievement tests.

6.2. Measures

6.2.1. Math and Dutch self-concept. Math and Dutch self-concept were assessed by means of six self-perceived competence items of the Self Description Questionnaire I (SDQ I; Marsh, 1992). The SDQ I is generally considered as one of the best empirically evaluated instruments for measuring preadolescents' self-concept (Byrne, 1996). From these six items, three items referred to students' mathematics self-concept (i.e., "I learn things quickly in

mathematics”; “I get good marks in mathematics”; “I’m good at mathematics”) whereas the other three items assessed their verbal self-concept (i.e., “I learn things quickly in Dutch”; “I get good marks in Dutch”; “I’m good at Dutch”). Items were scored on a five-point Likert-type rating scale ranging from 1 (*false*) to 5 (*true*). Reliabilities of both math self-concept and verbal self-concept were good (Cronbach’s alpha was .82 for both scales). For math self-concept and Dutch self-concept, percentages of missing data were 2.8% and 2.3%, respectively (with small inter-item variability).

6.2.2. Math and Dutch achievement. In order to measure students’ achievement in Dutch, a standardized spelling test (Scale Progress in Spelling V, SVS-V; Moelands & Rymenans, 2003) and an shortened version of a standardized reading comprehension test (CITO reading comprehension test; Staphorsius & Krom, 1998) were administered. To measure spelling, students were given a dictation consisting of monosyllabic and polysyllabic words wherein a number of specific aspects of Dutch spelling were tested (e.g., open versus closed syllable). In Grade 4, the more difficult B-version of the SVS-V was administered to all students. To measure reading comprehension, students were given a few short texts to read, after which they had to answer several multiple choice questions concerning this text. Students’ math achievement was measured by means of a curriculum relevant standardized multiple choice test covering five domains: mental arithmetic, arithmetic, algebraic problems, geometry, and applied arithmetic. For the reading comprehension and math test, an A-version and a B-version were administered to low- and high-performing students, respectively (this assignment was based on students’ test results in Grade 3). A calibration procedure using item response theory (IRT) techniques was administered in advance to enable comparison across test versions and time. Reliabilities of all achievement tests were good (Cronbach’s alphas ranging between .85 and .90). To create one global score for Dutch (language) achievement, we calculated the average of the spelling and reading comprehension test scores and

continued the analyses with this composite variable (if one of the two scores was missing, the value for the composite variable was also missing). Dutch achievement data were missing in 2% of the cases and math achievement data were missing in 1% of the cases.

6.3. Statistical analysis

Given that classes with complete missing data on one of the core variables were omitted in our final samples, percentages of missing data were low (overall lowest covariance coverage in the revisited model = .954; that is, at least 95% of the cases were available to build the covariances). In general, data were missing because some students were absent the day the tests and questionnaires were administered or because single items were left blank (in case of the ASC measures). In order to account for the missingness, Full Information Maximum Likelihood (FIML) estimation was used in Mplus 6.1 (Muthén & Muthén, 1998-2010). Additionally, the robust maximum likelihood (MLR) estimator in Mplus was used, providing standard errors and a chi-square (when applicable) robust to non-normality and non-independence (Muthén & Muthén, 1998-2007).

To investigate our hypotheses, we used a multilevel modeling approach as a meaningful way to take the hierarchical clustering of the data into account (students nested within classes) and to test the central predictions of the BFLP model (i.e., relation of individual and class-average achievement to students' ASC). Additionally, standard errors are more accurately estimated in multilevel models. Students nested within a higher level unit (classes or schools) tend to be more similar to each other than to students in other classes or schools. Statistically, this is likely to result in correlated residuals and in an underestimation of the standard errors. In general, the value added of multilevel models over more traditional approaches such as OLS regression has been widely established (Hox, 2010).

In contrast with previous I/E-BFLP studies that considered schools as the higher level unit (e.g., Chiu, 2012; Parker et al., 2013), the present study focused on classes rather than

schools. A substantial drawback associated with the clustering of students within schools is that the straightforward calculation of school level (Level 2) variables, as an aggregation of individual students (Level 1) scores, is likely to be affected by sampling bias because generally only a subset of Level 1 units is used. Complex latent aggregation approaches (Marsh et al., 2009) have been proposed to resolve this issue. However, because entire classes were considered in our study, we had measures available for almost every student in class. As a consequence, the sampling bias outlined above does not apply to the current research. In other words, given that only 1-2% of the Dutch and math achievement data were missing, we feel confident that we managed to calculate accurate class-level achievement variables with virtually no sampling bias, thus eliminating the need for latent aggregation. Additionally, we were not able to consider both the class and the school level in the context of the present study because many Flemish elementary schools only have one single class per grade. In accordance with findings of other studies (e.g., Marsh et al., 2005), ICC's for math and Dutch self-concept at the class level were small (ranging from .002 to .045).

Together with the multilevel approach, we used a latent modeling approach. The use of manifest variables in most multilevel models could be problematic in that most manifest variables are associated with measurement error, especially in social sciences (Parker et al., 2013). A failure to simultaneously control for measurement error hinders unbiased parameter estimation. More recent statistical packages such as Mplus version 6.1 (Muthén & Muthén, 1998-2010) enable researchers to estimate multilevel models while simultaneously controlling for measurement error (see Lüdtke et al., 2008 and Marsh et al. 2009 for more information on these kinds of models and their specifics). Standard fit indices were used to evaluate all models. For good fit, the chi square statistic should be as small as possible, the Root Mean Square Error of Approximation (RMSEA) less than or equal to .06, the Comparative Fit Index (CFA) and the Tucker-Lewis Index (TLI) equal to or larger than .95 (Hu & Bentler, 1999). As

the chi square statistic seems particularly sensitive to large sample sizes (Byrne, 2012; Vandenberg & Lance, 2000), we will only reject models with significant p -values for the chi square if the other fit indices indicate poor model fit as well. Nested models were compared using scaled chi-square difference tests.

Specifically, latent factors were created for math and Dutch self-concept; the three items for each self-concept dimension were used as indicators. To make sure that the same constructs were considered across the student and class level, corresponding factor loadings of latent factors were constrained to be equal across Levels 1 and 2 (Marsh et al., 2009). The measurement model encompassing both self-concept factors had a good fit, scaled $\chi^2(21) = 221.813$, $p < .001$, RMSEA = .047, CFI = .973, TLI = .962. We checked whether this model with constrained factor loadings fit the data significantly worse than the model with free factor loadings, but this was not the case, scaled χ^2 difference test (4) = 8.720, $p = .068$, $\Delta\text{CFI} < .010$. We did not control for measurement error in achievement, as we only had one score for each achievement test.

Finally, we standardized the individual achievement scores and all self-concept items across the sample mean (which is equivalent to grand-mean centering; Enders & Tofighi, 2007). Group-average achievement variables were not re-standardized after being calculated. Effect sizes (ES) were calculated for Level 2 effects using the formula mentioned by Seaton et al. (2011; i.e., $ES = 2 \times B \times SD_{\text{predictor}} / SD_{\text{criterion}}$), which is equivalent to Cohen's d (Cohen, 1988). Significance of these effect sizes was obtained by calculating these effect sizes in Mplus using the Model Constraint command (see Parker et al., 2013).

7. Results

7.1. Descriptive analyses

Table 1 presents all means, standard deviations, and intercorrelations for the manifest study variables. All achievement variables were positively correlated with each other and

individual achievement was positively correlated with ASC. In line with previous I/E findings, we found a small positive correlation between Math and Dutch self-concept scores, which was smaller than the correlation between math and Dutch achievement scores.

7.2. I/E Model

The I/E model showed good fit to the data, scaled $\chi^2 (29) = 219.100, p < .001$, RMSEA = .038, CFI = .983, TLI = .974). All hypothesized relations were confirmed (see Table 2). Strong within-domain regressions of self-concept on achievement were observed in the math and Dutch domain: higher math and Dutch achievement were associated with a more positive math and Dutch self-concept, respectively. Additionally, we found evidence for negative cross-domain relations between achievement and self-concept. Math achievement was negatively related to Dutch self-concept and Dutch achievement was negatively related to math self-concept. In conclusion, together with the low correlation between math and Dutch self-concept, our results lend further support for the I/E Model in elementary school with positive within-domain effects and negative cross-domain effects between math/ Dutch achievement and math/ Dutch self-concept (Hypothesis 1).

7.3. BFLP Model

Both, in the math domain, scaled $\chi^2 (6) = 11.510, p = .074$, RMSEA = .014, CFI = .999, TLI = .997) and the Dutch domain³, scaled $\chi^2 (7) = 30.426, p < .001$, RMSEA = .028, CFI = .994, TLI = .990, our models showed good fit to the data. At the student level, individual math achievement was positively related to math self-concept and individual Dutch achievement was positively related to Dutch self-concept. At the class level, our results were consistent with our hypotheses, confirming that class-average math and Dutch achievement were negatively related to students' math and Dutch self-concept, respectively (see Table 3; $ES_{\text{math}} = -.66, p < .001$; $ES_{\text{verbal}} = -.47, p < .001$ with $ES = 2 \times \text{unstandardized regression coefficient} \times SD_{\text{predictor}} / SD_{\text{criterion}}$ and absolute values larger than .20, .50, and .80

reflecting small, moderate, or large effects; Cohen, 1988). As such, our results confirm the hypothesis that for equally able students more able class environments lead to a lower academic self-concept both in the math and Dutch domain – even in Grade 4 (Hypothesis 2).

7.4. The Revisited I/E Model

Overall, the fit of the Revisited I/E Model was good, scaled $\chi^2 (37) = 242.921, p < .001$, RMSEA = .035, CFI = .983, TLI = .974. In line with our third main hypothesis, the central predictions of both the I/E Model and the BFLP Model seemed to hold when they were combined in a unifying model (see Table 4 and Figure 2). Of note, these findings indicate that both the I/E and the BFLP model explain independent parts in ASC variability. Moreover, the unexplained variance of Dutch and math self-concept at Level 2 decreased with 38% and 78% respectively by adding the class-level predictors to the I/E model. As such, the unified model adds incremental predictive validity to the explanation of variability in ASC in elementary school.

Additionally, we found some support for positive cross-domain compensatory effects of class-average achievement on ASC. We observed a positive effect of class-average Dutch achievement on math self-concept and of class-average math achievement on Dutch self-concept, suggesting that higher class-average levels of Dutch/math achievement were associated with a higher math/Dutch self-concept. However, the size of these positive compensatory effects should not be overestimated. Effect size was only small (ES = 0.14 for both effects) and a model constraining both compensatory paths to zero fit the data equally well: we only found a small difference in scaled $\chi^2 (2) = 14.13, p < .001$ which may be trivial considering our large sample size and $\Delta CFI < .01$ (Cheung & Rensvold, 2002; Vandenberg & Lance, 2000). Our results further largely support the domain specificity of the BFLP effect given that class-average achievement was negatively related to academic self-concept in the same domain only.

8. Discussion

Academic self-concept is important in understanding student learning and development even for elementary school students. Up to now, internal and external comparison processes influencing academic self-concept have scarcely been examined together in elementary education. The present study aimed to fill this research gap. Based on the predictions of the I/E Model and the BFLP Model, we tested a unifying model in which the effects of individual student achievement and class-average achievement in both the math and verbal domain were simultaneously modeled. Our findings demonstrate that the combination of the I/E and the BFLP model in one unifying model gives us a better understanding of elementary school students' ASC than each model alone because it adds incremental predictive validity to the explanation of variability in academic self-concepts in elementary school students. Even after controlling for the effects of internal comparison processes, external comparison processes were still related to academic self-concept and vice versa. In comparing the effect sizes of the results of the I/E Model and the BFLP Model separately with the effect sizes of the revisited model, I/E and BFLP effects were quite stable across models. This further supports the statement that both models uniquely predict academic self-concept variability. Therefore, their combination into one model leads to a better explanation – in terms of explained variability and parsimony – of academic self-concept formation than both single models.

Our results of the revisited model specifically showed that students' math achievement positively predicted math self-concept and negatively predicted their verbal (Dutch) self-concept. In line with the predictions of the I/E Model, a similar pattern was observed for the effects of verbal achievement (i.e., Dutch spelling and reading comprehension) on the verbal (Dutch) and math self-concept, respectively. The Revisited I/E Model further indicated that when class-average achievement in both domains was considered simultaneously, class-

average math and verbal achievement were strongly negatively related to academic self-concept within domains, as would be expected from within the BFLP Model. Additionally, we found some support for the existence of the positive cross-domain compensatory effects of group-average achievement that were observed by Parker et al. (2013) in themed secondary schools. Nevertheless, class-average verbal/math achievement only had a small beneficial effect on students' math/verbal self-concept in our study. Positive compensatory cross-domain effects of group-average achievement variables thus seem to be of larger importance when there is more differentiation between the achievement domains under consideration (the correlation between both group-average achievement variables for themed schools in the Parker et al. (2013) study was .39, whereas this correlation was .72 in our study).

Furthermore, the present study provides evidence that the Revisited I/E Model plays a crucial role in academic self-concept formation, not only in secondary education, but also in elementary education. In line with other I/E studies in elementary education (see Möller et al., 2009 for an overview), we found a small positive correlation between the math and verbal self-concept of Grade 4 elementary students. The small size suggests that elementary school students are already able to identify themselves in differentiated terms although there seems to be less differentiation than in older students. Parker and colleagues (2013), for instance, reported small negative correlations between the math and verbal self-concept of upper secondary students. Importantly, we found strong evidence for a BFLP effect of class-average achievement on academic self-concept in Grade 4 students. In line with previous BFLP studies with elementary school students and class as a frame of reference (e.g., Köller et al., 2008; Thijs et al., 2010), this suggests that social comparisons already play an important role in how younger students view themselves in the academic domain.

The present study also had some limitations which may inspire future research. First, our data were cross-sectional. The BFLP Model and the classic I/E Model are typically only

tested on the basis of a single cross-sectional wave of data, but the Reciprocal Effects Model (REM) describes how the associations between prior academic self-concept and subsequent achievement (and vice versa) longitudinally develop over time. A pertinent issue for further research is thus how to integrate the premises of the REM into the Revisited I/E Model. More recently, Möller, Retelsdorf, Köller and Marsh (2011) as well as Niepel, Brunner, and Preckel (2014) already showed that the central predictions of the classic I/E Model hold when longitudinally modeled. As students often compare themselves with different groups over time, complex cross-classification analytical approaches will be required or meaningful subgroups will need to be identified (i.e., students in stable versus changing compositions). Second, although we hypothesized that our results occurred because of internal and external comparison processes, we did not observe these processes directly in the present study. Future studies thus may benefit from using qualitative methods assessing more explicitly how students experience their academic context and how they construe their academic self-concept. Finally, we specifically considered the effect of class-average achievement in Flemish elementary schools. As such, more research is needed to establish whether our results may be replicated in other elementary school systems across different countries and cultures (and with, for example, other relevant reference frames).

Nevertheless, our study has some important implications. First, our findings imply that the Revisited I/E model for ASC formation encompassing both internal and external comparison processes should be favored instead of two separate models each focusing on one type of process. The present study thereby also highlights the need for further integration with other important frameworks, such as the REM (see above). Furthermore, our results suggest that teacher, educators, and other professionals in the school context should be made aware that even younger students in elementary schools are making both internal and external comparisons to form their ASC. Due to internal comparisons, students, who are exceptionally

good in math, may underestimate themselves in Dutch and vice versa. Due to external comparisons, equally able students may have different views of themselves as learners according to the achievement level of their class group. These effects may even be reinforced by the teacher who may also be influenced by the class-average achievement level (Baudson, Fischbach, & Preckel, 2015; Trautwein & Baeriswyl, 2007). In sum, this study extends research by empirically combining the I/E Model and the BFLP Model on academic self-concept formation in a sample of elementary school students. The revisited model shows more specifically that internal comparison processes, resulting in negative cross-domain effects of individual achievement on academic self-concept, go hand in hand with external comparison processes, resulting in positive effects of individual achievement and negative effects of class-average achievement within a domain on academic self-concept.

References

- Baudson, T. G., Fischbach, A., & Preckel, F. (2015). Teacher judgments as measures of children's cognitive ability: A multilevel analysis. *Learning and Individual Differences*. Advance online publication. doi: 10.1016/j.lindif.2014.06.001
- Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 15, 1–40.
doi:10.1023/a:1021302408382
- Brunner, M., Keller, U., Dierendonck, C., Reichert, M., Ugen, S., Fischbach, A., & Martin, R. (2010). The structure of academic self-concepts revisited: The nested Marsh/Shavelson model. *Journal of Educational Psychology*, 102, 964–981.
doi:10.1037/a0019644
- Byrne, B. M. (1996). *Measuring self-concept across the life span: Issues and instrumentation*. Washington, D.C.: American Psychological Association.
- Byrne, B. M. (2012). *Structural equation modeling with Mplus*. New York, NY: Taylor & Francis.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233–255. doi:10.1207/S15328007SEM0902_5
- Chiu, M. S. (2012). The internal/external frame of reference model, big-fish-little-pond effect, and combined model for mathematics and science. *Journal of Educational Psychology*, 104, 87–107. doi:10.1037/a0025734
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. New York: Academic Press.
- Dickhäuser, O., Reuter, M., & Rilling, C. (2005). Coursework selection: A frame of reference approach using structural equation modeling. *British Journal of Educational*

- Psychology*, 75, 673-688. doi:10.1348/000709905X37181
- Dijkstra, P., Kuyper, H., van der Werf, G., Buunk, A. P., & van der Zee, Y. G. (2008). Social comparison in the classroom: A review. *Review of Educational Research*, 78, 828-879. doi:10.3102/0034654308321210
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12, 121-138. doi:10.1037/1082-989x.12.2.121
- Ferla, J., Valcke, M., & Cai, Y. (2009). Academic self-efficacy and academic self-concept: Reconsidering structural relationships. *Learning and Individual Differences*, 19, 499-505. doi:10.1016/j.lindif.2009.05.004
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117-140. doi:10.1177/001872675400700202
- Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, 95, 124-136. doi:10.1037/0022-0663.95.1.124
- Hox, J. J. (2010). *Multilevel analysis. Techniques and applications (2nd Edition)*. New York: Routledge.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, 49, 505-528.
- Huguet, P., Dumas, F., Marsh, H. W., 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, 156–170. doi:10.1037/a0015558
- Köller, O., Zeinz, H., & Trautwein, U. (2008). Class-average achievement, marks, and academic self-concept in German primary schools. In H. Marsh, R. H. Craven & D. M. McInerney (Eds.), *Self-processes, Learning, and Enabling Human Potential: Dynamic New Approaches. Advances in Self Research, Vol. 3*, (pp. 331-352). Charlotte, NC: Information Age Publishing Inc.
- 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, 203-229.
- Maes, F., Ghesquière, P., Onghena, P., & Van Damme, J. (2002). *Longitudinaal onderzoek in het basisonderwijs: Van doelstellingen tot onderzoeksopzet* [Longitudinal Research in Primary Education. From objectives to research design]. Leuven, Belgium: Steunpunt Loopbanen doorheen Onderwijs naar Arbeidsmarkt.
- 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.
- Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, 23, 129-149.
- Marsh, H. W. (1990a). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology*, 82, 623– 636. doi:10.1037/0022-0663.82.4.623
- Marsh, H. W. (1990b). Influences of internal and external frames of reference on the formation of math and English self-concepts. *Journal of Educational Psychology*, 82, 107–116. doi:10.1037/0022-0663.82.1.107

- Marsh, H. W. (1992). *Self Description Questionnaire (SDQ) I: A theoretical and empirical basis for the measurement of multiple dimensions of preadolescent self-concept. An interim test manual and research monograph*. Macarthur, New South Wales, Australia: University of Western Sydney, Faculty of Education.
- Marsh, H. W. (1994). Using the national longitudinal study of 1988 to evaluate theoretical models of self-concept: The Self-Description Questionnaire. *Journal of Educational Psychology*, 86, 439–456. doi:10.1037/0022-0663.86.3.439
- Marsh, H. W., & Hau, K. (2003). Big-fish-little-pond effect on academic self-concept. A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*, 58, 364–376. doi:10.1037/0003-066X.58.5.364
- Marsh, H. W., & Hau, K. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology*, 96, 56-67. doi:10.1037/0022-0663.96.1.56
- Marsh, H. W., Köller, O., & Baumert, J. (2001). Reunification of East and West German school systems: Longitudinal multilevel modeling study of the big-fish-little-pond effect on academic self-concept. *American Educational Research Journal*, 38, 321-350. doi:10.3102/00028312038002321
- Marsh, H. W., Kong, C., & Hau, K. (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, 337–349. doi:10.1037/0022-3514.78.2.337
- Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling

error. *Multivariate Behavioral Research*, 44, 764-802.

doi:10.1080/00273170903333665

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement:

Relations and causal ordering. *British Journal of Educational Psychology*, 81, 59–77.

doi:10.1348/000709910x503501

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

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

Marsh, H. W., Trautwein, U., Lüdtke, O., Baumert, J., & Köller, O. (2007). The big-fish-little-pond effect: Persistent negative effects of selective high schools on self-concept after graduation. *American Educational Research Journal*, 44, 631–669. doi:10.3102/0002831207306728

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, 76, 397-416. doi: 10.1111/j.1467-8624.2005.00853.x

Moelands, F., & Rymenans, R. (2003). *Schaal Vorderingen in Spellingvaardigheid voor Vlaanderen (SVS-V)*. Handleiding. Arnhem: Citogroep.

Möller, J., & Marsh, H. W. (2013). Dimensional Comparison Theory. *Psychological Review*, 120, 544-560. doi:10.1037/a0032459

- Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research*, 79, 1129-1167.
doi:10.3102/0034654309337522
- Möller, J., Retelsdorf, J., Köller, O., & Marsh, H. W. (2011). The reciprocal internal/external frame of reference model an integration of models of relations between academic achievement and self-concept. *American Educational Research Journal*, 48, 1315-1346. doi:10.3102/0002831211419649
- Muthén, B. O., & Muthén, L. K. (1998-2010). *Mplus version 6.1 [statistical software]*. Los Angeles, CA : Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (1998-2007). *Mplus user's guide (5th ed.)*. Los Angeles, CA: Muthén & Muthén.
- Niepel, C., Brunner, M., & Preckel, F. (2014). The longitudinal interplay of students' academic self-concepts and achievements within and across domains: Replicating and extending the reciprocal internal/external frame of reference model. *Journal of Educational Psychology*. Advance online publication. doi:10.1037/a0036307
- 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.
doi:10.1016/j.learninstruc.2012.07.001
- Pinxten, M., Marsh, H. W., De Fraine, B., Van Den Noortgate, W., & Van Damme, J. (2014). Enjoying mathematics or feeling competent in mathematics? Reciprocal effects on mathematics achievement and perceived math effort expenditure. *British Journal of Educational Psychology*, 84, 152-174. doi:10.1111/bjep.12028

- Preckel, F., & Brüll, M. (2010). The benefit of being a big fish in a big pond: Contrast and assimilation effects on academic self-concept. *Learning and Individual Differences*, 20, 522–531. doi:10.1016/j.lindif.2009.12.007
- Preckel, F., Zeidner, M., Goetz, T., & Schleyer, E. J. (2008). Female ‘big fish’ swimming against the tide: The ‘big-fish-little-pond effect’ and gender-ratio in special gifted classes. *Contemporary Educational Psychology*, 33, 78–96.
doi:10.1016/j.cedpsych.2006.08.001
- Seaton, M., Marsh, H. W., & Craven, R. G. (2009). Earning its place as a pan-human theory: Universality of the big-fish-little-pond effect across 41 culturally and economically diverse countries. *Journal of Educational Psychology*, 101, 403–419.
doi:10.1037/a0013838
- Seaton, M., Marsh, H. W., & Craven, R. G. (2010). Big-fish-little-pond effect: Generalizability and moderation—two sides of the same coin. *American Educational Research Journal*, 47, 390–433. doi:10.3102/0002831209350493
- Skaalvik, E. M., & Rankin, R. J. (1995). A test of the Internal/External Frame of Reference Model at different levels of math and verbal self-perception. *American Educational Research Journal*, 32, 161–184. doi:10.3102/00028312032001161
- Skaalvik, E. M., & Skaalvik, S. (2002). Internal and external frames of references for academic self-concept. *Educational Psychologist*, 37, 233–244.
doi:10.1207/S15326985EP3704_3
- Staphorsius, G., & Krom, R. (1998). *Toetsen Begrijpend Lezen: Handleiding, opgavenboekjes en formulieren [Reading Comprehension Tests: Manual, test books, and forms]*. Arnhem: CITO.

- Thijs, J., Verkuyten, M., & Helmond, P. (2010). A further examination of the big-fish-little-pond effect: Perceived position in class, class size, and gender comparisons. *Sociology of Education*, 83, 333–345. doi:10.1177/0038040710383521
- Trautwein, U., & Baeriswyl, F. (2007). When high-achieving classmates put students at a disadvantage: Reference group effects at the transition to secondary schooling. *Zeitschrift fur Padagogische Psychologie*, 21, 119-133. doi:10.1024/1010-0652.21.2.119
- Vandenberg, R. J., & Lance, C. E. 2000. A Review and synthesis of the measurement invariance literature: Suggestions, practices and recommendations for organizational research. *Organizational Research Methods*, 3, 4-70. doi:10.1177/109442810031002
- Wouters, S., Colpin, H., Van Damme, J., De Laet, S., & Verschueren, K. (2013). Children's academic self-concept formation: Do classmates or friends matter most? *Learning and Individual Differences*, 27, 193-200. doi:10.1016/j.lindif.2013.09.002
- Wouters, S., Germeijs, V., Colpin, H., & Verschueren, K. (2011). Academic self-concept in high school: Predictors and effects on adjustment in higher education. *Scandinavian Journal of Psychology*, 52, 586–594. doi:10.1111/j.1467-9450.2011.00905.x

Footnotes

¹ Throughout this article we sometimes use the word “effect”. In contrast with self-concept theory which discusses the causal relationships between achievement and self-concept, we refer to predictive effects without implying causality in the current article.

² Because our cross-sectional sample was retrieved from a large-scale research project, we will shortly specify the global outline of the data selection procedure. In a first step, all Grade 4 students were selected with either academic self-concept data and/or achievement data, resulting in a sample of 4,978 students (nested in 299 classes). In a second step, classes were omitted wherein no self-concept questionnaires or achievement tests were administered. As a result of this, 17 classes ($N=254$) were no longer considered. In a last step, classes with less than 10 students were deleted, resulting in our final sample of 4,436 Grade 4 students (nested in 241 classes).

³ The BFLP model for the verbal domain has one degree of freedom more than the BFLP model for the math domain, because we had to fix one negative residual variance to zero for the latent verbal factor in the between part of the model. This constraint is present in all models including the verbal domain.

Table 1
Intercorrelations, Means, and Standard Deviations for all Variables

Variable	1	2	3	4	5	6
1. Individual math achievement	--					
2. Individual Dutch achievement	.60*	--				
3. Class-average math achievement	.52*	.35*	--			
4. Class-average Dutch achievement	.37*	.48*	.72*	--		
5. Math self-concept	.48*	.20*	.04	.03	--	
6. Dutch self-concept	.05*	.32*	-.02	.00	.10*	--
Raw <i>M</i>	96.48	79.08	96.46	79.05	3.72	3.69
Raw <i>SD</i>	8.57	7.90	4.42	3.84	1.01	0.95
<i>Min</i>	67.35	47.91	78.50	64.74	1.00	1.00
<i>Max</i>	116.68	109.41	108.27	88.69	5.00	5.00
<i>N</i>	4390	4358	4436	4436	4379	4379

Note. * $p < .001$. Taking our large sample size into account, we used $p < .01$ as the lower limit for significant correlations.

Table 2

Test of the I/E Model in Grade 4 (Model 1; N = 4433)

	Math self-concept	Dutch self-concept
Individual math achievement	.72 (.02)*	-.24 (.02)*
Individual Dutch achievement	-.16 (.02)*	.55 (.02)*

Note. Standardized estimates are reported. Standard errors in parentheses. Within-domain effects in bold. * $p < .001$.

Table 3

Multilevel Models for the BFLP effect on math self-concept (Model 2a; N = 4390) and verbal self-concept (Model 2b; N = 4358) in Grade 4

	Unstandardized estimates
Model 2a. Math self-concept	
Individual math achievement	.47 (.02)*
Class-average math achievement	-.42 (.03)*
Model 2b. Verbal self-concept	
Individual verbal achievement	.32 (.02)*
Class-average verbal achievement	-.33 (.04)*

Note. Standard errors in parentheses.* $p < .001$.

Table 4

Results for the Integration of the BFLP Model and the I/E Model into one Model (Model 3; N = 4436)

	Math self-concept	Dutch self-concept
Individual math achievement	.54 (.02)***	- .17 (.02)***
Class-average math achievement	- .47 (.03)***	.10 (.04)*
Individual Dutch achievement	- .11 (.02)***	.41 (.02)***
Class-average Dutch achievement	.10 (.04)**	- .36 (.05)***

Note. Unstandardized parameter estimates are reported. Standard errors in parentheses. Within-domain effects in bold. * $p < .05$; ** $p < .01$. *** $p < .001$.

(a) I/E model

(b) BFLP model (verbal)

(c) BFLP model (math)

(d) Revisited I/E model

Figure 1. Predictions for different models under consideration. MAch = math achievement, VAch = verbal achievement, MSc = math self-concept, VSc = verbal self-concept.

Figure 2. Integrated model representing the Revisited I/E Model in Grade 4. Effect sizes are reported for all class-average achievement effects on academic self-concept, otherwise standardized values are reported. MAch = math achievement, DAch = Dutch achievement, MSc = math self-concept, DSc = Dutch self-concept. * $p < .05$. ** $p < .01$. *** $p < .001$.