Differential school contextual effects for math and English: Integrating the big-fish-little-pond effect and the internal/external frame of reference

Philip D. Parker\textsuperscript{a,b,*}, Herbert W. Marsh\textsuperscript{b,c,d}, Oliver Lüdtke\textsuperscript{a}, Ulrich Trautwein\textsuperscript{a}

\textsuperscript{a} Center for Educational Science and Psychology, University of Tübingen, Germany
\textsuperscript{b} Centre for Positive Psychology and Education, University of Western Sydney, Australia
\textsuperscript{c} Department of Education, University of Oxford, United Kingdom
\textsuperscript{d} King Saud University, Saudi Arabia

\textbf{A R T I C L E   I N F O}

Article history:
Received 20 April 2011
Received in revised form 27 June 2012
Accepted 2 July 2012

Keywords:
Academic self-concept
Social comparison
Achievement
Multi-level modeling
School contexts

\textbf{A B S T R A C T}

The internal/external frame of reference and the big-fish-little-pond effect are two major models of academic self-concept formation which have considerable theoretical and empirical support. Integrating the domain specific and compensatory processes of the internal/external frame of reference model with the big-fish-little-pond effect suggests a potential unified model and the presence of a positive compensatory school context effect. This effect, in contrast to the big-fish-little-pond effect, indicates that school-average ability in one academic domain has a positive effect on self-concept in another domain. To date, little empirical support has been observed for this underlying theoretical assumption. The unified model and implied positive compensatory school context effect were tested in both traditional and themed (magnet) university track schools in Germany. In a large multilevel study (5016 students from 157 schools), we found support for the unified model in both school types but for the positive compensatory effect in themed schools only. The substantive and applied implications of these findings are discussed, as well as their theoretical importance for self-concept research.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Academic self-concept is a central area of interest to educational research, not only as an important outcome in itself, but also as an important predictor of other educational outcomes (Guay, Larose, & Boivin, 2004; Marsh, 1991, 2007). It is now well established that self-concept and achievement are reciprocally related over time (Marsh, 2007) and that self-concept is associated with academic effort, anxiety, confidence, motivation, and persistence in education (Marsh, 2007; Stankov, Lee, Lou, & Hogan, in press). A growing body of research also indicates that academic self-concept is an important predictor of coursework selection and other educational choices made in school and beyond (Marsh, 1991, 2007; Nagy et al., 2008; Parker et al., in press). Indeed, Eccles (2009) indicates that academic self-concept plays a central determining role in the educational and career choices that young people make. As such, the degree to which the school context facilitates or hinders the development of self-concept is an important educational concern.

Given its significance, considerable research has been conducted on self-concept development, with much of this research focused on the role of social comparisons and frames of reference (Marsh, 2007). Students spend much of their time in school, and the culture of most western schools is evaluative and competitive, where comparison with peers is a central feature (Covington, 2000; Liem, Givns, Martin, Stone, & Herrett, 2012). Thus, the school context is thought to play a particularly important role in academic and general self-concept formation (Dijkstra, Kuyper, van der Werf, Buunk, & van der Zee, 2008). The internal/external frame of reference and the big-fish-little-pond effect\textsuperscript{1} represent two theories that conceptualize the role of frames of reference and school context on self-concept, which have found widespread empirical support and which have important implications for educational policy (Marsh & Hau, 2003, 2004; Wong & Watkins, 2001). Despite their importance, both models have limitations, yet these may be overcome by integrating them into a single framework. Below we introduce the

\textsuperscript{1} We use the word effect throughout this article in keeping with the language of the big-fish-little-pond effect. It should be noted, however, that while self-concept theory discusses the causal relationships between achievement (at the individual and context level) and self-concept (see Marsh, 2007), in the context of the current research we refer to predictive effects rather than causal effects.
internal/external frame of reference model and the big-fish-little-pond effect, outline a proposed single unified model of self-concept formation, and explore the implications that a unified model may have for school policy, such as whether themed/magnet schools (schools whose selection policy targets ability in specific academic domains) create a school context that is beneficial for students’ self-concept development.

2. Internal/external frame of reference and self-concept formation

Academic self-concept refers to an individual’s relatively stable and organized perception and evaluation of themselves within socially important academic domains (Shavelson, Hubner, & Stanton, 1976). Academic self-concept factors are strongly related to academic achievement, where higher achievement in an academic subject is associated with more positive self-concepts in that domain. Indeed, achievement and self-concept are reciprocally related, such that an increase in achievement predicts more positive self-concepts but as academic self-concepts become more positive, achievement also increases (Marsh, 2007). Achievement, however, appears to be a necessary but not sufficient predictor of academic self-concept. Rather, extensive research has uncovered the role of social context in the formation of academic self-concept beyond that explained by objective achievement alone (see Ireson & Hallam, 2009; Marsh, 2007). This research has been largely driven by applying findings from social psychology, including research on social comparisons and frames of reference, to self-beliefs in education contexts (see Dijkstra et al., 2008; Suls & Wheeler, 2000 for reviews). The major finding of this research is that individuals’ perceptions of themselves are not solely objective but are also influenced by the context within which they evaluate themselves.

Marsh’s (1986) internal/external frame of reference is a model of self-concept formation that differentiates outside influences from internal comparison processes. The external dimension stresses the influence of external performance cues in the development of a student’s academic self-concept. This external frame of reference refers to the information available in their environment that individuals use as a basis for assessing their own competence, including the perceived abilities of others and external achievement feedback (e.g., feedback from experts). The theory notes that the most salient frame of reference for children is their school and thus their assessments of themselves within academic domains are, at least in part, dependent on their own self-perceived ability compared with that of their school peers and teacher feedback. The internal dimension refers to comparisons that students make across disciplines in assessing their own strengths and weaknesses (see Eccles, 2009). There is increasing recognition in educational research that students hold multidimensional perceptions about their schooling, which lead to internal evaluations and comparison (Hofer, Kuhnhle, Killian, Marta, & Fries, 2011). Indeed, the internal frame of reference was stimulated by findings that academic self-concept consists of two near orthogonal factors: math (consisting of sub-factors such as math, sciences, and economics self-concepts) and verbal (consisting of sub-factors such as geography, history, foreign language, and native language self-concepts; see Marsh, 2007 for reviews). The lack of relationship between these two self-concepts is surprising, given that achievement in these domains is positively correlated. Marsh theorized that this makes sense, however, if students form their self-concept not only by comparing their performance in a subject domain against external performance feedback but also by internally comparing their performance in different domains against each other. This internal comparison is ipsative, such that an increase in self-concept in one domain (i.e., verbal or math) triggers a decrease in self-concept in another domain (Marsh, 2007). Bringing together both external and internal frames of reference, Marsh predicted that student achievement would be positively related to self-concept if both were in the same academic domain but negatively related to self-concept in other domains (see Fig. 1: Model 1).

The internal/external frame of reference (hereafter I/E model) has received widespread cross-cultural support (Marsh & Hau, 2004). In reviewing the literature, Möller, Pohlmann, Kölker, and Marsh (2009) note the multiplicity of divergent sources of evidence in favor of this model. First, there is evidence that the expected relationships emerge when frames of reference are experimentally manipulated (e.g. Möller & Kölker, 2001). Second, a number of longitudinal studies have consistently shown that increases in achievement over time predict a more positive self-concept in the same domain but have negative implications for self-concept in other domains, as predicted by the I/E model. Such research also points to the critical role that I/E processes play in achievement noting that positive self-concept in one academic domain predicts increases in achievement in the same domain and small decreases in achievement in other domains (see Möller, Retelsdorf, Kölker, & Marsh, 2011). Third, studies analyzing the contents of students’ introspective diaries indicate students do make both external and internal comparisons in their daily life, and both comparison processes inform their self-concept (Möller & Husemann, 2006). Fourth, the meta-analysis conducted by Möller et al. (2009) indicated that associations between achievement and self-concept predicted by the I/E model were present across different measures, countries, age groups, and gender. Finally, research by Nagy et al. (2008), Marsh (2007), and Parker et al. (in press) have shown that self-concept ipsatively predicts domain specific educational choices in patterns predicted by the I/E model. For instance, students who have a high self-concept in math tend to avoid choosing disciplines such as history or social sciences.

3. Big-fish-little-pond effect and self-concept formation

In a classic study, Davis (1966) found that college men’s achievement was positively related to entry into elite careers. For students of equal ability, however, the more able their peers the less likely they were to enter an elite profession. This led Davis (1966) to conclude “the aphorism ‘It is better to be a big frog in a small pond than a small frog in a big pond’ is not perfect advice, but it is not trivial” (p. 31). Marsh (1991) built on this work, as well as theoretical developments in sociology and psychology, to explain why students in high achieving schools had lower academic self-concepts than would be expected on the basis of the their objective achievement alone (Marsh, 1991). Marsh (1991) labeled this phenomenon the big-fish-little-pond effect (hereafter BFLEP).

The BFLEP model suggests the counterfactual case that a student placed in a high-ability school will have lower self-concept than if they had been placed in a lower-ability school. This is because the average ability of the student peers will differ depending on the school they attend and thus their relative ability rank within each school will also be different (Trautwein, Gerlach, & Lüdtke, 2008). Formulated as a set of hypotheses, the BFLEP predicts that school-average ability will be negatively related to students’ academic self-concept after controlling for individual achievement (see Fig. 1: Models 2a & 2b). This model has extensive cross-cultural support (Marsh & Hau, 2003) and quasi-experimental support (Marsh, Kölker, & Baumert, 2001; Wouters, De Fruinane, Colpin, Van Damme, & Verschueren, in press), as well as strong support in a variety of academic domains (e.g., Dijkstra et al., 2008; Trautwein et al., 2008).
concept has suggested that schools with higher average ability may be associated with a 'reflected glory' effect for students' self-concept (see Dijkstra et al., 2008). This theory states that a student's self-concept will be positively affected by moving to a prestigious school because the student's acceptance into the school is perceived to reflect positively upon their own abilities. The empirical literature, however, indicates that the positive benefit to self-concept of attending a more prestigious school is weaker than the BFLPE and that the overall relationship between school average achievement and student self-concept is negative (Dijkstra et al., 2008).

The BFLPE effect is not merely an expression of students' internal lives; rather comparison with peers is an integral part of most school systems, where evaluations about students are often made in comparison to other students (see Covington, 2000 for a review). For example, Espenshade, Hale, and Chung (2005) observed that elite college entry boards make decisions on potential students, at least in part, on the average ability of the school they attended. BFLPE processes are thus not naive but, at least to some degree, reflect an external reality. Furthermore, the BFLPE has important implications not just for the individual student but also for educational policy. In particular, the BFLPE suggests that school selection policies like selective schooling or the mainstreaming of children with a disability may be counter-productive as they place children in high average ability contexts and thus may have a negative effect on academic self-concept. Thus, research is needed which explores how school context and educational policies affect academic self-concept.

4. Integrating the I/E and BFLPE models

Both the I/E model and the BFLPE are empirically supported models for understanding academic self-concept formation, however, both models have limitations. While the I/E model provides a rich framework for understanding domain specificity in academic self-concept, it is a student level model and does not take into account contextual factors (i.e., it does not detail how students' academic self-concepts are affected by the ability composition of their school). The vast majority of research in this area uses single level regression models, which do not include any school level variables. In contrast the BFLPE is a contextual effects model that does not take into account the multi-dimensional nature of academic self-concept. Thus, the majority of research in this area uses either domain general academic measures or focuses only on a single academic domain.

Skaalvik and Skaalvik (2002) suggest similar limitations, noting that research on the I/E model provides relatively little detail on how students' academic self-beliefs are affected by the ability composition of the school they are in. Skaalvik and Skaalvik (2002), further note that one of the major outstanding questions of academic self-concept research is the extent to which students' internal frames of reference are affected by the ability composition of their school, when multiple academic domains are explored. This is indeed problematic as the extreme domain specificity of academic self-concept suggests that, in many cases, theoretical models, which do not take the multidimensional nature of self-
concept into account are likely to be inappropriate (Marsh, 2007). Thus an important goal in furthering research on academic self-concept is to identify whether the BFLPE and the I/E models can be integrated to form a single unified model to account for their respective shortcomings.

By integrating the I/E and BFLPE models two hypotheses can be drawn. First, the domain specificity of self-concept suggests the negative relationship between school average ability and student self-concept will only be negative when both are measured in the same academic domain (Skaalvik & Skaalvik, 2002). Second, the internal frame of reference in the I/E model indicates that anything that lowers self-concept in one domain should result in a more positive evaluation of self-concept in other academic domains. Thus, the relationship between school average ability and self-concept in one academic domain would be expected to have a positive compensatory effect on self-concept in other academic domains. Put simply, if a student feels like their classmates are better than them in math, they may feel more negative about their math ability but this may result in them evaluating their English ability more positively. This is because an internal frame of reference is based on the use of perceived achievement in one academic domain as a reference point for assessing achievement in other areas (Marsh & Hau, 2004). Thus, anything that results in a decline in self-concept in one domain (e.g., domain specific BFLPE) should theoretically have positive implications for self-concept in other domains (Marsh, 1986).

5. Existing research findings on integrating the I/E and BFLPE models

In one of the few studies to incorporate both the I/E and BFLPE model predictions, Marsh (1986) focused on the relations between math and English achievement and self-concept as a test of the domain specificity of the BFLPE. In addition to providing support for the unique importance of both the BFLPE and the internal/external frame of reference models, evidence in favor of the domain specificity of the BFLPE and a positive compensatory effect of school-average ability in one domain (e.g., math or verbal) leading to more positive self-concept in the non-matching domain was observed. However, the support for the positive compensatory effect in this study was weak. This may have been because the research was not conducted under conditions that were most conducive for testing this effect. In part this was due to latent variable and multilevel statistical modeling (and their integration) not being well-developed nor widely available at that period (see online methodology for more detail). However, limitations of this research were not only of a methodological nature. In particular, Marsh (1991) suggested that the positive compensatory effect was small because school average achievement in math and English were highly correlated. This context is the case in most schools, as selection procedures tend to lead to “school contexts of math and of English achievements [being] very similar” (p. 114). Thus, potential positive compensatory effects are likely to be obscured. However, selection processes that either explicitly or implicitly target specific academic domains (or themes) would likely lead to more differentiated school contexts for different academic domains, providing a prime context for testing the positive compensatory effect. Thus, support for a positive compensatory effect suggests that high ability schools with a differentiated math and verbal ability contexts may mitigate some of the negative effects of the BFLPE.

6. Traditional and themed (magnet) schools

This differentiation across academic domain is likely to be apparent in targeted selective schools called themed (or magnet) schools. These schools are different from traditional university track (selective) schools in that they have a specialized set of programs focused on an academic domain or on preparing students for a post-school university major or career in a particular field (such schools are often referred to as magnet schools in the United States and Australia; Gamoran, 1996; Giffin, Allen, Kimura-Walsh & Yamamura, 2007). In such schools, students are either selected or self-selected based primarily on their interests or achievement in a narrowly defined area of excellence (e.g., math or literature ability). In contrast, traditional university track schools are selective in terms of general academic excellence (e.g., math and literature ability). Thus, when compared to traditional university track schools who attract students of high ability across academic domains, students in themed schools are likely to be similar in the academic domain most closely associated with the focus of the school but more variable in other domains. As such, school contexts in math and verbal domains will likely be more differentiated in themed rather than traditional schools and thus well suited to testing the unified model and the implied positive compensatory effect developed in this research.

7. The current research

The purpose of the current research was to test a unified model of academic self-concept formation, which integrates the internal/external frame of reference and the BFLPE. In order to do this we tested a set of three hypotheses, which are outlined in detail below and are summarized in Table 1. First, in the introduction we suggested that the positive compensatory effect would emerge in themed selective schools rather than in traditional selective schools, because school average ability is likely to be more differentiated. Thus, as a prerequisite to testing the unified model, we hypothesized that the correlation between school average achievement in math and English would be significantly smaller in themed schools (i.e., more differentiated) than in traditional

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Hypotheses.</th>
<th>Supported if:</th>
<th>Fig. 1:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Themed schools will be more differentiated in achievement than general schools</td>
<td>Correlation between school average math and English ability is smaller in themed school than in general schools</td>
<td>Model 2a</td>
</tr>
<tr>
<td>2a</td>
<td>UE model will be valid for both school types</td>
<td>Model provides a good fit to the data and parameter estimates are consistent with theory</td>
<td>Model 3a</td>
</tr>
<tr>
<td>2b</td>
<td>BFLPE for math will be valid for both school types</td>
<td>Model provides a good fit to the data and parameter estimates are consistent with theory</td>
<td>Model 3b</td>
</tr>
<tr>
<td>2c</td>
<td>BFLPE for English will be valid for both school types</td>
<td>Model provides a good fit to the data and parameter estimates are consistent with theory</td>
<td>Model 3c</td>
</tr>
<tr>
<td>3a</td>
<td>Unified model of self-concept formation will be valid for both school types</td>
<td>School average achievement in one domain will only have a negative effect on self-concept in the same domain</td>
<td>Model 3</td>
</tr>
<tr>
<td>3b</td>
<td>BFLPE will be domain specific</td>
<td>Parameters from school average ability in one domain significantly and positively predict self-concept in another domain</td>
<td>Model 3</td>
</tr>
<tr>
<td>3c</td>
<td>Positive compensatory effect will be significant for themed schools only</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
schools. Second, we aimed to confirm that both the BFLPE and the internal/external frame of reference models were relevant to students in both traditional and themed university track schools. While both models have been well validated in a variety of settings, little research has applied either model to themed or magnet schools specifically. In both cases, we expect these patterns of associations to be supported, regardless of school type, and thus to represent general models of academic self-concept formation.

Finally, we aimed to test a unique model of self-concept formation and in particular how the school context can help or hinder self-concept development. As the BFLPE, I/E, and unified models posit results that are seemingly counter-intuitive, it is critical that teachers understand how they work. In addition, the unified model has important implications for themed schools as an educational policy. Given the narrowly defined area of excellence of themed schools, the prediction derived from the unified model is that themed schools are likely to undermine self-concepts in the domain that is their focus. However, in such schools the positive compensatory effect is expected to be significant. This would serve, at least partially, to compensate for the negative effect of school average achievement on self-concept in one academic domain by creating conditions in which self-concept in other domains might be boosted.

8. Method

8.1. Participants

The current research utilized data from the ongoing Transformation of the Secondary School System and Academic Careers (TOSCA) project. This project follows two cohorts of students during the transition from high school. The survey responses from these two cohorts represent some of the most important databases in Germany for testing hypotheses about academic achievement, motivation, and self-concept in young adults. The two cohorts of the study have contributed significantly to the self-concept literature. For example, research with the first cohort of the project indicated that the BFLPE is a persistent influence on self-concept even after high-school graduation (Marsh, Trautwein, Lüdtke, Baumert, & Köller, 2007). Research with the second cohort has indicated that both BFLPE and reflected glory are present in German university track schools and represent significant influences on students’ self-concept formation (Trautwein, Lüdtke, Marsh, & Nagy, 2009). Several other studies focusing on the role of self-concept in predicting high-school graduation destinations and the relationship between the BFLPE and personality are also currently being undertaken (see Trautwein, Neumann, Nagy, Lüdtke, & Maaz, 2010 for an overview of the TOSCA project).

The data for the current research comes from the second cohort of the TOSCA project, which began in 2006. This ongoing project focuses on university track upper secondary schools, where data was obtained from students in 157 randomly selected upper secondary schools in a single German state. Participants filled out a questionnaire that included the analysis variables used in this research during a specially assigned in-class session conducted during school hours. They were given the option of not participating in the research, yet approximately 80% of the potential sample, or a total of 5016 participants, took part in the research. The research was conducted when participants were in their final year of upper secondary schooling. The average age of the participants...
was 19.76 ($SD = 1.12$). The sample was weighted toward females (58 percent), which is consistent with the demographic makeup of German university track schools.

8.2. Traditional and themed university track schools in Germany

The university track in Germany (Gymnasium) consists of approximately 35% of the secondary school student population (Eurydice, 2010). It is an important context for the current research, as the system includes clearly demarcated traditional and themed university track schools. The Gymnasiale oberstufe (upper high schools) system includes both large numbers of traditional upper high schools (Allgemeinbildendes Gymnasium; 58% of participants in the current research) and a number of selective themed upper high schools (Berufliches Gymnasium), which students can attend from the age of 15 (Eurydice, 2010). Both school types are pre-university schools, meaning that most graduates will enter university after having finished school (see Buchmann & Dalton, 2002; Dustmann, 2004; Hillmert & Jacob, 2010 for details). The themed schools in Germany largely consist of students who have come from vocational school tracks, where average achievement is generally lower than for traditional schools, and they often consist of a more disadvantaged population (Köller et al., 2004). Nevertheless, students from both school types aspire to get the Abitur, a school-leaving certificate that provides the broadest access to higher education (Eurydice, 2010). In the current research, themed schools included those that focused on business management, technical/engineering, bio-technology, agricultural sciences, nutrition sciences, and social pedagogy themed schools. While these themes differ in their focus, they are similar in being orientated toward some academic domains more than others.

Themed schools students had lower achievement scores in math ($d = -0.73$, $p < .001$) and English ($d = -0.58$, $p < .001$) than traditional school students. A moderate difference was also observed for English self-concept ($d = -0.23$, $p < .001$) but not for math self-concept ($d = -0.05$, ns). Consistently with the focus of themed schools on disadvantaged youth, students from these schools were more likely to be born overseas ($p = .06$, $p < .001$) and to have a mother ($p = .05$, $p < .01$) and/or a father ($p = .06$, $p < .001$) who spoke a language other than German at home. Parents of themed school students also had lower levels of education ($d = -0.46$, $p < .001$), and worked in jobs with lower occupation prestige ($d = -1.17$, $p < .001$) than traditional school students. This is consistent with themed schools being an alternative path to university, particularly catering for disadvantaged students.

8.3. Measures

8.3.1. Achievement

To measure academic achievement, we utilized well-validated and comprehensive standardized tests of math and English. The mathematics achievement test that was administered was taken from the Third International Mathematics and Science Study (TIMSS; e.g., Baumert, Bos, & Lehmann, 2000). The test was administered with a matrix design where students worked on only a subset of the item pool (Nagy & Neu mann, 2010). The math test consisted of three components, including calculus, geometry, and equations. English achievement was assessed using a shortened research version of the Test of English as a Foreign Language (TOEFL), as used in the Institutional Testing Program (Educational Testing Service, 1997). This instrument comprises three components: listening comprehension, structure and written expression, and reading comprehension. Internal consistency estimates (Cronbach’s alphas) were high, with estimates of .93 for math achievement and .97 for English achievement. School-average achievements in English and math were measured by aggregating individual student achievement (see Table 3 for inter-class correlations).

8.3.2. Self-concept

Math and English self-concepts were measured using the German version of the SDQ III (Schwanzer, Trautwein, Lüdtke, & Sydow, 2005). The SDQ III is a multidimensional self-concept instrument for late adolescents and young adults and includes a number of domain specific factors based on the Shavelson et al. (1976) model. Previous research with the SDQ instruments indicates its excellent construct validity and reliability in German (Schwanzer et al., 2005). From the 17 scales in the SDQ III (German), only the math self-concept factor (e.g. I have never been good at tasks that required mathematics [reverse scored]) and the English (as a second language) self-concept factor (e.g. I am just not that talented in English [reverse scored]), added in recent versions of the German SDQ III, were used. Participants responded to each item on a 4-point (agree-disagree) response scale. Internal consistency estimates for self-concept were high, with estimates of .90 for math self-concept and .92 for English self-concept.

8.4. Statistical analysis

8.4.1. Multi-level modeling

The current research utilized multilevel modeling to test the major hypotheses. Multilevel modeling focuses on treating nested levels (e.g. students nested within schools) within data as meaningful. This approach has two major advantages. First it allows researchers to test hypothesized relationships at different levels in the data simultaneously. For example, how are student achievement and school average achievement associated with self-concept? Second, individuals nested within a higher level (e.g. students in the same schools) are likely to be more similar to each other than with members nested within different, higher level units (e.g. students in other schools). Thus, residuals are likely to be correlated and traditional analysis approaches (e.g. OLS regression) will tend to underestimate standard errors. Multilevel modeling provides a means for calculating correct standard errors and thus appropriate tests of statistical inference (see Hox, 2010 for a general introduction to multilevel modeling and estimation procedures). Multilevel modeling is well-established in the social sciences and has been used frequently in BFLPE research (e.g. Marsh et al., 2009). Nevertheless, readers may be unfamiliar with three additional components of our approach which are not typical of multilevel modeling. We thus provide a detailed discussion of these elements.

First, contextual effects like the BFLPE have typically been tested by looking at whether the composition of the school (e.g. school

| Table 3 |
|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | Math Ach | English Ach | Math SC | English SC |
| General         |           |             |        |           |
| Math achievement| .03       | .74         | .92    | .38      |
| English achievement| .42      | .01         | .83    | .68      |
| Math self-concept| .79       | .20         | .05    | .46      |
| English self-concept| -.03      | .65         | -.16   | .01      |
| Themed          |           |             |        |           |
| Math achievement| .10       | .39         | .78    | -.54     |
| English achievement| .36      | .04         | .52    | .33      |
| Math self-concept| .69       | -.03        | .04    | -.31     |
| English self-concept| -.04      | .67         | -.26   | .01      |

Notes. Average inter-class correlations in bold on the diagonals. Correlations in the upper quadrant represent relationships at the school level. Correlations in the lower quadrant represent relationships at the participant level. All relationships are significant at $p < .05$. |
average achievement) predicts self-concept over and above student achievement. In order to test this, we first group mean-centered (where group equals the schools) student achievement. This allows for the direct computation of within and between-group estimates. The school context effect is then given by subtracting the within-group estimate from the between-group estimate (Raudenbush & Bryk, 2002). With group-mean centering, this is simply the difference between the coefficient for the school average achievement and the coefficient for individual achievement (see the online Appendix A for how to estimate this with correct standard errors). If school context makes a difference (i.e. contributes to self-concept net of individual achievement), then this compositional effect will be significantly different from zero, and the direction of the effect will indicate the nature of the relationship (i.e. is it a positive or negative predictor; see Harker & Tynms, 2004 for an extensive review). Effect size estimation for contextual effects is not yet well developed (see Marsh et al., 2009 for a review). Thus, we used two different effect size measures; one (ES1) based on the residual variance at Level 1 (e.g. the participant level) and one based on the total Level 1 variance (ES2). The difference between these two approaches is that ES2 typically is more conservative. For further discussion see the Appendix A.

Second, most multilevel models use manifest variables such as scale scores. This can be problematic when such variables are measured with error (the typical case in social sciences) and thus, parameter estimates are likely to be attenuated (Cohen, Cohen, Aiken, & West, 2003). The use of latent variables, where the measurement error is modeled and parameter estimates are controlled for this error is an important advance over the use of manifest variables. Recent advancements, and the wide availability of statistical packages that incorporate latent variables and multilevel data (e.g. Mplus), allow for the inclusion of latent variables in multilevel models (see Preacher, Zyphur, & Zhang, 2010 for a review). We take advantage of this ability in order to provide less-biased parameter estimates.

Finally, multilevel modeling can be affected by sampling bias when aggregation used to create Level 2 (e.g. school level) variables is based on only a subset of Level 1 units. Put simply, aggregated school average achievement may be biased if data is available for only a subset of the students within each school. This was relevant to the current research, in which an average of 30 students were sampled from each school. To overcome this, we used latent aggregation (see Marsh et al., 2009). In this approach the available units (i.e. the students for whom we do have data) are treated as indicators of Level 2 variables (e.g. school average achievement). Analogously to the latent variables described above, this allows a means of directly modeling and thus controlling for, sampling error. When latent variable analysis is used in conjunction with latent aggregation, as is done in this research, multilevel models control for two sources of error; measurement and sampling. Such models are referred to as doubly latent (Marsh et al., 2009).

8.4.2. Missing data and multi-group analysis

To test the hypothesis that school achievement contexts in math and English would be more differentiated in themed schools (Hypothesis 1), and that the proposed compensatory effect would be apparent in themed but not in traditional schools (Hypothesis 3), we used a multi-group approach to data analysis. In this case, a model where parameters of interest for both school types were constrained to be equal was first estimated. The fit of this model was then compared with a model in which parameters of interest to testing the hypotheses were free to take on different values for each school type. A chi-squared difference test was used to compare the fit of these models. In relation to missing data, which was small (<5%), a full-information-maximum-likelihood (or FIPLM) approach was used. FIPLM is useful, as it combines missing data and parameter estimation in a single step and uses all the available information (Enders, 2010).

9. Results

The three proposed hypotheses were tested sequentially using different sets of latent multi-level models. The hypothesis that math and English school average achievement are more differentiated in themed than in traditional schools was tested in a multi-level confirmatory factor analysis using a multi-group framework (see Appendix A for a description of multi-level CFA models). The second hypothesis was that both the I/E (Fig. 1: Model 1) and BFLPE models (Fig. 1: Model 2a and 2b) could be replicated in both school types. These models were submitted to the data using a multi-group approach. The final hypothesis was that a unified model would fit the data well and that a positive compensatory effect of school average achievement in one domain on self-concept in another domain would be significant. This hypothesis was tested by submitting the unified model of self-concept formation to the data using a multi-group approach (Fig. 1: Model 3).

9.1. Hypothesis 1: differential relationships for themed and traditional schools

Hypothesis 1 was tested using a multilevel confirmatory factor analysis (hereafter MCFAs; see Mehta & Neale, 2005). MCFAs test the adequacy of hypothesized measurement structures purported to underlie the data at both student and school level (see Appendix A). Using a multi-group approach we first fitted a constrained model in which all parameters were held invariant across school type. The constrained model fitted the data well, indicating good construct validity of the instruments used to measure self-concept and achievement at both student and school level ($$\chi^2 = 338$$; $${RMSEA} = .03$$). A model in which latent correlations between the achievement and self-concept were free to vary across school types however, provided a significantly better fitting model than the constrained model (see Table 2). This suggested that there were different patterns of relationships between achievement and self-concept in math and English in themed versus traditional schools (Table 3). Supporting Hypothesis 1, the relationship between school average math and English achievement was $$r = .74$$ in the traditional schools but only $$r = .39$$ in the themed schools ($$p < .001$$). This indicates that math and English school average achievement were significantly more differentiated in themed than in traditional university track schools and that themed schools provided the conditions likely to give rise to a positive compensatory effect—should such an effect exist.

9.2. Hypothesis 2: validity of the I/E and BFLPE models across school type

9.2.1. I/E model

An I/E model was submitted to the data using a multi-group approach to test the hypothesis that the associations predicted by this model were significant in both themed and traditional schools. The I/E model is an individual level model, and thus the focus was on relationships between achievement and self-concept at the student level, with standard errors corrected for the nested nature of the data (see Muthén & Satorra, 1995). Confirming that the I/E model was valid across school types, a constrained model provided a good fit to the data ($$\chi^2 = 40.236$$; $${CFI} = .97$$; $${TLI} = .97$$; $${RMSEA} = .05$$) and supported expected relationships. A free model was also fitted in order to see if there were any differences between the school types in the strength of the I/E relationships. This model
Table 4
I/E model results.

<table>
<thead>
<tr>
<th></th>
<th>General</th>
<th>Themed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Math SC</td>
<td>Eng SC</td>
</tr>
<tr>
<td>Math Ach</td>
<td>1.14 (.06)**</td>
<td>-.42 (.03)**</td>
</tr>
<tr>
<td>Eng Ach</td>
<td>-.21 (.03)**</td>
<td>.82 (.03)**</td>
</tr>
</tbody>
</table>


9.2.2. BFLPE

Models testing the BFLPE effect for the math (Model 2a) and English (Model 2b) domains separately were submitted to the data using a multi-group approach. The constrained model provided an excellent fit for the math model ($\chi^2$ [79] = 391; CFI = .98; TLI = .98; RMSEA = .03) and the English model ($\chi^2$ [77] = 250; CFI = .99; TLI = .99; RMSEA = .03). Models were then fitted to the data in which the parameter estimates were free to vary across school types. The chi-squared difference test for both Models 2a and 2b indicated that the free models provided significantly better fit to the data than the constrained model (see Table 2), indicating that some differences between the school types were present. As with the I/E model these differences reflected the strength of effects, rather than differences in the pattern of results. For math (Model 2a) the negative BFLPE was slightly weaker in the traditional (B = -.60, p < .001, ES1 = -.39, ES2 = -.24) than the themed (B = -.41, p < .001, ES1 = -.41, ES2 = -.30) schools. Likewise, for the English model the negative BFLPE was weaker in the traditional schools (B = -.41, p < .001, ES1 = -.18, ES2 = -.11) than in the themed (B = -.67, p < .001, ES1 = -.35, ES2 = -.23) schools. Taken together, however, these results were consistent with the hypotheses in indicating that higher school-average ability had a negative effect on student academic self-concept and that this contextual effect was present in both math and verbal domains, regardless of school type (Table 5).

9.3. Hypothesis 3: testing an integrative model

Analyses then tested a unified model of self-concept formation which integrated the I/E framework and the BFLPE (Fig. 1: Model 3). This integrative model was submitted to the data, where it provided a good fit to the data ($\chi^2$ [340] = 1051; CFI = .97; TLI = .97; RMSEA = .04). While the constrained model provided a good fit to the data, a model in which hypothesized parameter estimates of interest were free to take on different values in traditional and themed schools provided a significantly better fit to the data (see Table 2). The difference between the traditional and the themed school in relation to the traditional I/E and BFLPE effects was in relation to the size of the effects rather than the pattern of results, where results generally supported the hypotheses drawn from both theories, regardless of school type (see Table 6). Importantly, school average ability was only a negative predictor of student self-concept in the same domain—indicating that the BFLPE is domain specific. This was the case for both school types. However, a difference in the pattern of results was observed for the effects of school average ability in one domain on self-concept in another domain. This relationship supported the proposed compensatory effect in themed schools for school average math achievement on English self-concept (B = .27, p < .01, ES1 = .15, ES2 = .10) and school average English achievement on math self-concept (B = .73, p < .01, ES1 = .86, ES2 = .47). Consistently with Hypothesis 3, this compensatory effect was smaller and not statistically significant in the traditional schools (see Table 6 and Fig. 2).

10. Discussion

The current research found support for all three hypotheses (see Table 1). First, a multi-level CFA indicated that math and English school contexts were more differentiated in themed than in traditional schools. Second, while small differences in the size of the parameter estimates were observed across school type, the I/E model and the BFLPE were supported in both themed and traditional schools. Finally, the I/E and the BFLPE were integrated into a single unified model of self-concept formation, which was submitted to the data. This model fitted the data well and significant positive compensatory relationships between school average ability in one domain and self-concept in another domain were observed in the themed schools but not in the traditional schools. The following discussion considers the implications of these results for theory and practice.

Table 5
BFLPE: Results for math and English domain.

<table>
<thead>
<tr>
<th></th>
<th>General</th>
<th>Themed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Math SC</td>
<td>Eng SC</td>
</tr>
<tr>
<td>Math Ach</td>
<td>1.01 (.04)**</td>
<td>.68 (.02)**</td>
</tr>
<tr>
<td>School average math Ach</td>
<td>-.88 (.06)**</td>
<td>-.76 (.03)**</td>
</tr>
</tbody>
</table>


Table 6
BFLPE-CE model results.

<table>
<thead>
<tr>
<th></th>
<th>General</th>
<th>Themed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Math SC</td>
<td>Eng SC</td>
</tr>
<tr>
<td>Participant level</td>
<td>Math Ach</td>
<td>1.11 (.05)**</td>
</tr>
<tr>
<td></td>
<td>Eng Ach</td>
<td>-.22 (.03)**</td>
</tr>
<tr>
<td>School level</td>
<td>Math Ach</td>
<td>-.81 (.23)**</td>
</tr>
<tr>
<td></td>
<td>Eng Ach</td>
<td>-.37 (.23)</td>
</tr>
</tbody>
</table>

10.1. Integration of the I/E and BFLPE models of academic self-concept

Supporting our hypothesis, a good fit to the data was observed for the unified model of academic self-concept formation and parameter estimates were consistent with hypotheses. In particular, the results supported the domain specificity of the BFLPE. Such a result is important as it suggests that, when subject domain is pertinent to the research question or context being addressed in research, the BFLPE should be pursued from a domain specific rather than from a domain general perspective. In addition, evidence indicated that school average ability in different academic domains was more differentiated in themed than in traditional schools. It is suggested here that this differentiation provided the foundation for the significant effect from school average ability in one domain predicting more positive self-concept in another domain. These results indicate that self-concept is influenced by a complex set of psychological processes which are affected in multiple ways by the context in which students’ self-evaluations take place (Skaalvik & Skaalvik, 2002).

10.2. Implications of the positive compensatory effect

The positive compensatory effect is not just important in supporting a unified model of self-concept formation but also has substantive implications in its own right. Previous research has indicated that the composition of a school has a number of implications for students’ educational outcomes (Harker & Tymms, 2004). For self-concept, school composition has typically been researched in relation to two competing effects; the BFLPE and the reflected glory effect (Dijkstra et al., 2008). Such research has generally suggested that the negative BFLPE exerts a stronger influence on students’ self-concept than the positive effect of being a member of a high achieving group (i.e., reflected glory). The current research, however, reveals a more complex, but also more positive, picture of the role of school composition on students’ academic self-concepts. In particular, schools selection policies with a focus on a narrow area of excellence may undermine self-concept in the domain that is their focus. However, unlike selection policies that focus on general academic excellence, themed school selection policies may also create conditions in which this...
negative effect is partially compensated for by boosting self-concept in domains that are not the focus of the school. It is well known from educational motivation research that school policy and classroom culture affects students’ self-beliefs (see Covington, 2000 for a review). Our research builds on this literature by suggesting that school selection policies also play an integral role in the intrapsychic life of students.

10.3. Practical implications

Themed (or magnet) schools in both the United States and Germany are important in providing an alternative to university and as a means to promote university entry for disadvantaged youth (Giffin et al., 2007; Köller et al., 2004). Thus, the ability of these schools to promote psychological factors that are known to predict educational choices and university entry, is important. This point is particularly the case for self-concept, which has such a central role in predicting the educational and occupational choices that young people make (Eccles, 2009). In the current research a moderate BFLPE was observed in themed schools. This raises important questions about the efficacy of themed schools as an educational policy. From this perspective these schools aim to provide a supportive context for students, to keep them in school and guide them toward university. To achieve this, the selection procedures used in such schools target students who have high achievement or interest in the school’s focus. The BFLPE suggests that the nature of such a selection policy may have a negative effect on self-concept in the very domain that is the focus of the school. Thus, these selection procedures may be counterproductive in that they may have a negative effect on a variable known to be important for university entry and career pathways (Eccles, 2009; Marsh, 1991).

Indeed, BFLPE research has similar implications for other educational policies. For example, research has raised important questions about whether selective schooling and mainstreaming of children with a disability may be counterproductive, given the BFLPE research has indicated that such education policies have negative effect on students’ self-concept (Skaalvik & Skaalvik, 2002). In contrast to concerns about selective schooling and mainstreaming, however, the current research suggests some important caveats for themed schools where the applied implications are not necessarily negative. First, without such institutions, themed school students would either attend vocational track schools, where it is more difficult to qualify for university (see Maaz, Trautwein, Lüdtke, & Baumert, 2008), or they would attend traditional university track schools. Given that students in themed schools have, on average, lower achievement than their traditional school peers, an intuitive hypothesis suggests that they would have greater exposure to a more uniformly negative BFLPE in traditional schools, which have high average ability across academic domains. Second, it is important to note that a compensatory effect of school average ability was observed in themed schools but not in traditional schools. This suggests that students in themed schools are exposed to a potential compensatory self-concept process of comparable strength to the BFLPE that was not present in traditional schools. While testing these applied implications was clearly outside the scope of this research, the support for the compensatory effect observed here provides one metric from which the efficacy of themed schools can be assessed.

10.4. Strengths, limitations and directions for further research

The current research was focused on testing a set of hypotheses that were derived from considering the integration of the internal/external frame of reference model and the big-fish-little-pond model. Strong evidence for a unified model was observed and substantive and applied implications were identified. Notwithstanding these strengths, several limitations of the study should be noted to guide interpretation of the results and to stimulate further research. First, while the current sample provided an important context for the current research, given that cross-cultural findings indicate that German students show among the highest BFLPE and I/E relationships (Marsh & Hau, 2003, 2004), the generalizability of the findings to other cultures needs to be considered. This is particularly the case in countries such as the United States, in which themed (or magnet) schools are also part of educational policy. Second, in the current research, English achievement and self-concept represented students’ proficiency and self-beliefs in a second language. Typically, the I/E model is tested using beliefs about native language proficiency. Nevertheless, Möller, Strebel, Pohlmann, and Köller (2006) have shown that the I/E model operates similarly for German youth whether German or English domains are compared to math self-concept. The I/E model theoretically applies to all components of verbal self-concept, including second languages, but also to other subjects, including geography, philosophy, and history. While the current research using a second language self-concept strongly supported the I/E model, further research replicating both the I/E and the BFLPE model is needed here with other components of verbal self-concept is needed. This limitation also applies to math self-concept where, for example, replication of the current results using science or engineering self-concept is needed. Third, although students came from different school types (themed or traditional) they were all still in highly selective university tracks, which account for approximately the top third highest achieving students in the German secondary school population (Eurydice, 2010; Hillmert & Jacob, 2010). While the results derived from the I/E and BFLPE models in this research were consistent with those observed in broader populations (see Marsh & Hau, 2003, 2004), the results of the unified model may have been different if a representative sample of the entire German secondary school system had been used. Fourth, the ability context of this study focused on schools. Most research on the BFLPE is done with school as the context of interest, and this choice of ability context is particularly pertinent to the German system. Nevertheless, recent research suggests that a local dominance effect may be important for academic social comparisons. In particular, the relationship between average ability and self-concept may be stronger, the more proximal the context (Allick, Zell, & Bloom, 2010). Hence, analysis using school classes (where appropriate; see Footnote 2) or friendship group (where students choose their own comparison targets) represents an important extension of this research and of self-concept research in general. Finally, tests of causal relations predicted by the I/E, BFLPE, and the unified models

2 The level at which social comparison research occurs is an important concern in research. Research on the BFLPE typically has occurred at broad group level and thus differs from social psychological research, in which social comparison targets are often chosen by the participant and consist of a single individual. The reason for this is that the BFLPE is typically concerned with the effect of universal (e.g. school-wide) comparison standards rather than more specific standards (Dijkstra et al., 2008). Within BFLPE literature there is further differentiation between research that focuses on schools as the level of comparison and research that focuses on the classroom or stream. While the majority of research has been conducted at the school level, in some contexts class is more appropriate (e.g. Wong & Watkins, 2001). In this respect, the appropriate level at which to conduct research is dependent on the context in which the data was collected. In Germany, ability stratification occurs at the school level, as compared to many other countries, in which stratification occurs within schools. Furthermore, students in upper secondary schools in Germany do not undertake classes with the same group of peers across all subjects. Thus, in the current German context, school rather than class provides the best level at which to conduct research.
were based on responses from a single wave of data. Thus, while the BFLPE and the I/E models have received causal support in previous research (see Marsh, 2007), there is a need to consider their integration, as outlined by the unified model, from a longitudinal perspective.

10.5. Conclusions

The current research tested a unified model of self-concept formation, bringing the BFLPE and I/E frameworks together into a single model. The model was submitted to data from a large sample of German university track students, utilizing recent advances in the testing of multilevel contextual effects. Findings suggested that themed schools were more differentiated in average ability in different subject domains than traditional schools. Further, strong support for the I/E model at the individual level, and domain specific negative relationships between school-average ability and students’ level of self-concept (BFLPE) was found regardless of school type. The results support a single unified model, which integrates the I/E and BFLPE models and fits the data well. A positive compensatory effect that was theoretically implied by this unified model was significant in themed but not in traditional schools. Taken together, these results provide important initial evidence in favor of a unified model of academic self-concept as a parsimonious framework for future research. The current findings also have applied implications, including providing one potential metric for evaluating themed schools as an educational policy.

Acknowledgments

We would like to thank the Jacob foundation’s PATHWAYS to Adulthood program and the Ministry of Science, Research, and the Arts Baden-Württemberg, which funded this research. In addition we would like to thank members of the PATHWAYS to Adulthood program, Felix Thoemmes, Marjorie Seaton, Arief Liem, and Susan Colmar for their helpful comments on the manuscript.

Appendix A. Supplementary material

Supplementary data related to this article can be found online at http://dx.doi.org/10.1016/j.learninstruc.2012.07.001.

References


