

The Big-fish–little-pond-effect Stands Up to Critical Scrutiny: Implications for Theory, Methodology, and Future Research

Herbert W. Marsh · Marjorie Seaton ·
Ulrich Trautwein · Oliver Lüdtke · K. T. Hau ·
Alison J. O'Mara · Rhonda G. Craven

Published online: 9 May 2008
© The Author(s) 2008

Abstract The big-fish–little-pond effect (BFLPE) predicts that equally able students have lower academic self-concepts (ASCs) when attending schools where the average ability levels of classmates is high, and higher ASCs when attending schools where the school-average ability is low. BFLPE findings are remarkably robust, generalizing over a wide variety of different individual student and contextual level characteristics, settings, countries, long-term follow-ups, and research designs. Because of the importance of ASC in predicting future achievement, coursework selection, and educational attainment, the results have important implications for the way in which schools are organized (e.g., tracking, ability grouping, academically selective schools, and gifted education programs). In response to Dai and Rinn (*Educ. Psychol. Rev.*, 2008), we summarize the theoretical model underlying the BFLPE, minimal conditions for testing the BFLPE, support for its robust generalizability, its relation to social comparison theory, and recent research extending previous implications, demonstrating that the BFLPE stands up to scrutiny.

Quotations (associated page numbers) to the Dai and Rinn (2008) article are based on a prepublication version of the article available to the authors of this article that may have changed during the final preparation for publication.

The authors would also like to express thanks to David Dai and Anne Rinn for their encouragement and assistance to us in preparation of our article, whilst still acknowledging that they might not agree with all the views expressed here.

H. W. Marsh (✉) · A. J. O'Mara
Department of Education, University of Oxford, 15 Norham Gardens, Oxford OX2 6PY, UK
e-mail: herb.marsh@education.ox.ac.uk

M. Seaton · R. G. Craven
University of Western Sydney, Sydney, Australia

U. Trautwein · O. Lüdtke
Center for Educational Research, Max Planck Institute for Human Development, Berlin, Germany

K. T. Hau
The Chinese University of Hong Kong, Shatin, Hong Kong

Keywords Big-fish–little-pond-effect · Social comparison theory · Ability grouping · Academic self-concept

In its simplest form, the big-fish–little-pond effect (BFLPE) predicts that students have lower academic self-concepts (ASC) when attending schools where the average ability levels of other students is high compared to equally able students attending schools where the school-average ability is low. Findings support the BFLPE and are remarkably robust, generalizing over a wide variety of different individual student and contextual level characteristics, settings, countries, longterm followups, and research designs. The results also have important policy implications for the ways in which schools are organized (e.g., ability grouping, tracking, selective schools, gifted education programs, etc.).

An Advance Organizer

The Dai and Rinn (2008) critique of the BFLPE is reminiscent in many ways of the Dai (2004) comment on a cross-cultural study on the BFLPE by Marsh and Hau (2003). To differing extents, both critiques by Dai (2004) and by Dai and Rinn (2008) suggested that: the BFLPE might be a short-term, ephemeral effect; noted situations in which there might be positive effects of school-average ability on ASC that is claimed to be inconsistent with BFLPE predictions; argued that there might be a number of individual student or contextual characteristics that moderate the BFLPE; contrasted BFLPE research in educational settings with social psychological research based on social comparison theory (SCT); and emphasized seemingly contradictory evidence from gifted education research. We believe that in both reviews the authors painted an overly simplistic—in some cases erroneous—picture of some of the theoretical issues that have been addressed in BFLPE research and then critiqued our research in relation to their limited interpretation of our research. Hence, in this article we highlight theoretical issues and research findings that counter Dai and Rinn’s interpretations of our work and their conclusions. However, it is important to also note that there are many areas in which we agree with their suggestions, including the need for further research and some of their suggestions for more fruitful directions that this research might take. In fact, in this article we summarize some of our current “in press” and ongoing research in which we have already started addressing some areas that Dai and Rinn emphasized as limitations in existing research and important directions for future research. More generally it is hoped that a vigorous and vibrant exchange of ideas such as presented by Dai and Rinn (2008; Dai 2004) and our response will strengthen BFLPE research.

Dai’s (2004) critique centered on Marsh and Hau’s (2003) analysis of a large, multinational database in relation to the BFLPE. Marsh and Hau found strong support for the BFLPE and its cross-cultural generalizability for responses from nationally representative samples of approximately 4,000 15-year olds from each of 26 countries ($N=103,558$) included in the 2000 OECD Programme for International Student Assessment (PISA) study. In relation to educational research, this level of cross-cultural generalizability is remarkably strong. In response to the Dai (2004) review, Marsh *et al.* (2004) demonstrated that:

- Coupled with research reviewed by Marsh and Hau (2003), there is extremely strong support for internal validity, external validity, generalizability, and policy-practice implications of the BFLPE (p. 269).
- Cultural differences (e.g., conceptions of ability, collectivist vs. individualistic) might be expected to influence the size of the BFLPE. However, the consistent support for the cross-cultural generalizability of the BFLPE from our OECD PISA study,

coupled with its generalizability in relation to diverse groups and settings reviewed by Marsh and Hau (2003), demonstrates that the BFLPE is extraordinarily robust.

In a subsequent analysis of PISA 2003 that included an even more diverse set of 41 countries than the 26 countries in PISA 2000 considered by Marsh and Hau (2003), Seaton (2007 also see Seaton *et al.* 2008a, b) replicated support for the cross-cultural generalizability of the BFLPE, demonstrated that the BFLPE generalized across collectivist and individualist cultures and across economically developing and developed nations, and that the BFLPE effect size (0.49) was sufficiently large to warrant practical attention as well as being substantively and theoretically important. On the basis of their critique of Dai (2004), Marsh *et al.* (2004; also see Marsh 2005a, b) concluded that “The BFLPE stands up to critical scrutiny” (p. 269)—a conclusion that is not challenged by a critical reading of the subsequent Dai and Rinn (2008) review.

The intent of this review is to provide an overview of theoretical, methodological, and policy-related issues arising from BFLPE research, with a particular emphasis on concerns raised by the Dai and Rinn (2008) critique. We begin with an overview of the BFLPE paradigm, its theoretical basis, and the minimum requirements for testing it. Next we compare and contrast these aspects of the BFLPE based largely on educational psychology research with relevant SCT research based largely on social psychology research, as SCT seems to provide an alternative perspective to the BFLPE in terms of theory, methods, and empirical findings. Whereas the focus of the BFLPE is on ASC, it is also relevant to evaluate the implications for academic achievement and performance—particularly in the context of tracking, ability grouping, and special provision for gifted education. From here we move to a discussion of individual and school characteristics that are potential moderators of the BFLPE and the policy implications of this research. Finally, we conclude with a brief summary of some ongoing statistical and methodological issues in BFLPE research, progress in addressing these issues, and directions for future research.

The BFLPE Paradigm

Theoretical basis

The fundamental theoretical premise underlying the BFLPE is that perceptions of the self cannot be adequately understood if the role of frames of reference is ignored. The same objective characteristics and accomplishments can lead to disparate self-concepts depending on the frames of reference or standards of comparison that individuals use to evaluate themselves, and these self-beliefs have important implications for future choices, performance, and behaviors (Marsh 2007; Marsh and Craven 2006). Hence, psychologists from the time of William James (1890/1963) have recognized that objective accomplishments are evaluated in relation to frames of reference, noting that “we have the paradox of a man shamed to death because he is only the second pugilist or the second oarsman in the world” (p. 310). Historically, the theoretical underpinnings of frame-of-reference research contributing to the BFLPE derive from research on adaptation level (e.g., Helson 1964), psychophysical judgment (Marsh 1974; Parducci 1995; Parducci *et al.* 1969; Rogers 1941; Wedell and Parducci 2000), social psychology (Morse and Gergen 1970; Sherif 1935; Sherif and Sherif 1969; Upshaw 1969; Volkman 1951), sociology (Alwin and Otto 1977; Hyman 1942; Meyer 1970), social comparison theory (Festinger 1954; Diener and Fujita 1997; Suls 1977; Suls and Wheeler 2000), and the theory of relative deprivation (Davis 1966; Stouffer *et al.* 1949).

On the basis of this broad theoretical perspective (particularly that based on frame of reference effects; e.g., Marsh 1974), Marsh (1984a, b, 1990; Marsh and Parker 1984) formulated a theoretical model of the BFLPE as applied to ASC in an educational psychology setting. Assume that three students (X, Y, and Z) vary in terms of their objective academic ability relative to the entire population of students across all schools: X (slightly below-average ability), Y (average ability), and Z (slightly above-average ability). Although student Y has an average academic ability relative to the population of all students, if Y attends a high-ability school (i.e., a school where the school-average ability is above the average across all schools), Y would have an academic ability below the average ability level of other students in the school. This is predicted to result in Y having a below-average ASC. However, if Y attends a low-ability school (i.e., a school where the school-average ability is below the average across all schools), then Y would be above the average ability level in this school, leading to an above average ASC. In a similar vein, the ASCs of students X and Z will depend (positively) upon their objective academic abilities, but will also vary (negatively) with the school-average ability. According to this model, a given academic ability level leads to a distribution of psychological impressions, indicating that other constructs (and random error) also affect this mapping. Although there was support for such a model based on psychophysical research dating back to the early 1900s that was the primary basis of this early research (see Marsh 1974), Marsh (1984a; see also Schwarzer *et al.* 1982) specifically developed the BFLPE paradigm to understand the formation of ASC in school settings.

Although not a main focus of the present investigation, the growing support for the multidimensionality of self-concept and theoretical models positing self-concept as a multidimensional, hierarchical construct (see Marsh 1990, 2007; Marsh and Craven 2006; Marsh *et al.* 1983) are important in tests of the BFLPE. Historically, self-concept researchers emphasized a global, relatively undifferentiated measure of self-concept, also referred to as self-esteem. However, particularly in educational psychological research, many important academic outcomes are systematically related to ASC but relatively unrelated to self-esteem. General academic self-concept refers to students' self-perceptions of their academic accomplishments, their academic competence, their expectations of academic success and failure, and academic self-beliefs. Importantly, this general ASC can also be broken into components related to broad academic disciplines (e.g., math and verbal self-concepts) as well as even more specific components of academic self-concept related to specific school subjects (e.g., history, English, foreign language, mathematics, computer studies, science, etc.; see Marsh 2007). Early BFLPE research (Marsh and Parker 1984; Marsh 1987) demonstrated that support for the BFLPE was highly domain specific; whilst ASC was strongly influenced by individual student achievement (positively) and school-average achievement (negatively), neither individual nor school-average achievement had much effect on either global self-esteem or non-academic components of self-concept. This support for the domain specificity of the BFLPE provided strong support for the importance of a multidimensional perspective of self-concept in educational psychology research, but also supported the construct validity of interpretations of the BFLPE.

A series of predictions—some of which appeared to be paradoxical at the time they were first proposed (Marsh 1984a)—can be generated from this model. In particular the model predicts that:

1. ASC will be positively related to academic ability;
2. school-average ASC will be similar in high-ability and low-ability schools even though the corresponding ability levels of individual students are substantially higher in high-

- ability schools and substantially lower in low-ability schools (i.e., the frame of reference is largely established by the student's own school);
3. school-average ability will be negatively related to ASC after controlling for individual student ability;
 4. ASC will be more highly correlated with individual ability after controlling for school-average ability;
 5. ASC can be more accurately predicted from individual and school-level ability than from either of these predictors considered separately;
 6. the negative effect of school-average academic ability is specific to ASC and is unlikely to generalize to non-academic components of self-concept (e.g., physical self-concept); and
 7. because the frame-of-reference is established by school-average ability, all students in a high-ability school are predicted to have lower ASCs than would the same students if they attended a low-ability school; interactions between school-average and individual ability on ASC are expected to be small or non-significant.

Expanding on this theoretical model, Marsh (1987, 1990, 1991; Marsh and Craven 2002) posited that the BFLPE represents the net effect of a stronger negative BFLPE (a contrast effect) and a weaker positive (assimilation or reflected glory) effect. Although reflected-glory assimilation effects have a clear theoretical basis, these effects have been largely implicit and elusive in BFLPE studies. Marsh *et al.* (2000; also see Trautwein *et al.* 2008) addressed this issue in a large representative sample of Hong Kong high school students by specifically asking students to evaluate the pride that they felt in attending their high school. As previously found in BFLPE studies, higher school-average achievement led to lower ASC in their longitudinal study. However, they also found that higher perceived school status had a counter-balancing positive effect on self-concept (an assimilation effect) that they likened to reflected glory and feelings of pride in belonging to a high-achieving school. The net effect of these counterbalancing influences was clearly negative, indicating that the contrast effect was stronger than the assimilation effect. Attending a school where school-average achievement is high simultaneously results in a more demanding basis of comparison for students within the school to compare their own accomplishments (the negative contrast effects) and a source of pride for students within the school (the positive reflected glory, assimilation effects). Although theoretically important, the assimilation effect found in this study has been elusive in other research and not nearly as robust as the typical contrast effects found in other BFLPE research.

Placing the BFLPE in its broader historical context, the effect is a specific example of more general frame-of-reference effects that have been studied in psychology (see Sherif and Sherif 1969). In demonstrations of the BFLPE, Marsh (1984a) operationalized the standard of comparison to be the school-average ability level. This is consistent with more general models of frame-of-reference effects in psychophysics (Helson 1964) and social psychology (Upshaw 1969) even though more complicated models have been posited (Marsh 1974, 1983, 1984a; Marsh and Pardo 1978; Upshaw 1969). This model, of course, ignores the frame-of-reference established by classes within schools—particularly in schools with ability streaming such that there might be competing frames of reference due to the school (school-average ability) and the class (class-average ability). Although it is easy to generalize the model to classes instead of schools, the model makes no predictions about the relative importance of the school and the class (but see Trautwein *et al.* 2008).

Importantly, the model does not posit that individual and school-average ability are the only determinants of ASC or that there are no other individual student and contextual

characteristics that moderate the size—and possibly even the direction—of the BFLPE. Indeed, because reducing or even eliminating the negative consequences of the BFLPE has important theoretical and practical implications, many BFLPE studies have looked for moderators of the BFLPE (Lüdtke *et al.* 2005; Marsh 1987, 1991, 2007; Marsh *et al.* 1995, 2000, 2001; Marsh and Craven 2002).

Minimum methodological requirements for BFLPE studies

In the last two decades, multilevel modeling has become one of the central research methods for applied researchers in the social sciences and has had a profound effect on BFLPE research. A major advantage of multilevel modeling over single level analysis lies in the possibility of exploring relationships among variables located at different levels simultaneously (Goldstein 2003; Raudenbush and Bryk 2002; Snijders and Bosker 1999). In the typical application of multilevel modeling, outcome variables are related to several predictor variables at the individual level (e.g., students) and at the group level (e.g., classes, schools). In this literature, models that include the same variable at both the individual level and the aggregated group level are called contextual analysis models (Boyd and Iverson 1979; Firebaugh 1978; Raudenbush and Bryk 2002). The central question in such contextual studies is whether the aggregated group characteristic has an effect on the outcome variables after controlling for variables at the individual level. In contextual studies the critical question is the relative sizes of effects of individual and group-average constructs in predicting relevant outcome measures when both individual and group-average variables are included in the analysis. In this respect the BFLPE paradigm is a classic contextual study in which individual (level 1 = L1) and school-average (level 2 = L2) achievement are used to predict ASC and the appropriate statistical analysis involves multilevel modeling. Interestingly, in this general contextual study paradigm, there is no assumption that individuals actually compare themselves to others in their group, although such social comparison processes are a central feature of SCT studies that have also had an important influence on BFLPE research.

It is useful to present a clear statement about the minimal conditions needed to test the BFLPE (Lüdtke *et al.* 2005; Marsh 2007; Marsh and Craven 2002; Marsh *et al.* 2004; Seaton *et al.* 2008c). As emphasized in contextual effect research more generally (Boyd and Iverson 1979), the BFLPE is inherently a multilevel phenomenon that incorporates both the individual level (e.g., student) and group level (school or classroom). As in other contextual effect studies, the critical issue is whether the group level effect (school-average ability) has a significant effect after controlling for the corresponding and appropriate individual effects (individual ability). Hence, the minimal conditions to test the BFLPE are:

- a multilevel design with many schools and a substantial (representative or total) sample of students from each school;
- an objective measure of achievement for each individual student that is directly comparable over different schools and an appropriate measure of ASC; and
- tests of the effects of school-average achievement on ASC after controlling for the effects of individual student achievement.

Reviewing the BFLPE literature: Testing the theoretical basis and meeting the minimum methodological requirements

When conducting a “critical review” of the research literature, a key step in the review process is to establish the research questions that the review aims to elucidate. The studies

included in the review should only be studies that explicitly address that research question. Establishing the research question, aims, and selection criteria of the studies a priori can reduce bias in the review process (Torgerson 2003). Given that the BFLPE is a hypothesized relation between academic self-concept, individual ability (or achievement), and school average ability (or achievement), any research study included in a review of the BFLPE should minimally contain analyses of these three variables. Using the criteria listed above for BFLPE studies, many of the studies considered by Dai and Rinn (2008, Appendix A) are not tests of the BFLPE. For example, Butzer and Kuiper (2006) considered neither achievement nor ASC; Cheng and Lam (2007) neither controlled for prior ability nor examined class-average ability; and Stapel and Koomen (2001) did not consider achievement. Additionally, Huguét *et al.* (2001) and Blanton *et al.* (1999) did not consider school-average ability but subsequent reanalysis of both these studies by Seaton *et al.* (2008c) found clear support for the BFLPE (for further discussion see “Integration of BFLPE and SCT Paradigms” in relation to this study). Indeed, arguably half of the 26 studies included in Dai and Rinn’s Appendix do not address the BFLPE at all.

In some cases, Dai and Rinn (2008) acknowledge that they were drawing conclusions from a test of “a variant of the BFLPE (p. 7)” rather than the BFLPE itself. Whereas some of these variants of the BFLPE might be heuristic in terms of how to extend BFLPE research, it must be emphasized that most of these studies do not provide tests of the BFLPE and thus cannot be said to contradict, limit, or constrain support for the BFLPE. Indeed, we suggest that a more appropriate method for their review would be to present two separate, explicit reviews: one addressing the research question “is there evidence for a big-fish-little-pond effect?” (based on only studies that meet the criteria of a BFLPE study) and the other addressing the research question “what evidence from studies of related models (e.g., SCT, social identity theory) could be integrated into the BFLPE model to improve our understanding of self-concept processes?” Our reading of Dai and Rinn is that there is clear support for the BFLPE based on BFLPE studies, but that other research suggests ways in which the BFLPE could be expanded. Confusing these two research questions by not making this distinction clear undermines the potential value of findings based on each question considered separately.

Juxtaposing BFLPE and Social Comparison Theory (SCT)

Dai and Rinn (2008) seem to imply that the only—or at least the primary—theoretical basis for the BFLPE was SCT, based on the work by Festinger (1954) and more recent extensions of SCT. From this perspective, in the section entitled *Why Is the BFLPE Paradigm Flawed?*, Dai and Rinn claim that “The BFLPE is based on the assumption that people compare themselves externally only with a local norm in their immediate environment (e.g., school average)” (p. 19–20) and that “unless the social or cultural norms are extremely compelling, overwhelming individuals’ flexible choice, people will selectively use varied comparison criteria under different circumstances that better serve self-evaluation purposes” (p. 20). Whereas these assumptions are central issues in SCT, BFLPE research makes no assumptions that the school- or class-average ability is the only basis of comparison, nor does it assume that students do not use other, more varied comparison criteria like those considered in SCT. More generally, Dai and Rinn are critical of BFLPE because the “BFLPE research program has had minimal contact with the social comparison literature” (p. 14), but actually this has been an important direction of recent BFLPE research that we summarize here.

In response to Dai and Rinn (2008) it is important to emphasize that although the theoretical models underlying the BFLPE and SCT share many historical influences and

their juxtaposition has much to offer, the theoretical basis for the BFLPE is much broader than and distinct from SCT, and is based on a well-established body of research (see earlier discussion) that in some cases predates Festinger (1954) and in some cases does not assume that students actively compare their ability levels with others. Although the formation of a frame of reference that is central to the BFLPE might involve an active process of comparison with other students, it could also be based on information such as a distribution of grades or test scores provided by a teacher that would not necessarily involve any interaction with other students. From this perspective it is relevant to juxtapose these two theoretical approaches and research findings based on each—with particular emphasis on recent research that has attempted to integrate the two approaches.

Role of a generalized other

According to the BFLPE, students use the average level of academic accomplishments of other students within their school to form a frame of reference against which to evaluate their own academic accomplishments. In this sense, the comparison is imposed, implicit, in relation to a generalized other, and reflects a classic contextual or frame-of-reference effect. In marked contrast, much social psychology research is based on a very different paradigm stemming from SCT—what Dai and Rinn refer to as a self-engendered comparison. In this traditional SCT choice paradigm (hereafter we refer to this as the SCT paradigm although we recognize that this is not the only research strategy used by SCT researchers and that there are variations in how this paradigm has been applied), participants are asked to select a target person as a basis of comparison. In this sense the selection of the comparison person is explicit and resulting social comparison effects are evaluated in relation to a specific target person. Obviously, students have considerably more flexibility in choosing target persons in this SCT choice paradigm than in the BFLPE paradigm. Although it might be possible to characterize this SCT research as a contextual effect or frame-of-reference study in which the context is based on a single student, contextual or frame-of-reference effects and multilevel modeling perspectives that have been so important in BFLPE studies have been largely ignored in SCT (see Seaton *et al.* 2008c).

Surprisingly, this historically important construct of generalized other has had little emphasis in recent SCT research that has focused more on variations of the traditional SCT choice paradigm: the choice of specific target individuals as a basis of comparison, the juxtaposition of upward and downward comparison strategies, and how these strategies satisfy competing needs. Indeed, because Festinger's early research that is the basis of SCT emphasized group processes, he also emphasized this notion of a generalized other as a basis of normative comparisons between the self and a group—how individuals use groups to evaluate their abilities and opinions (Suls and Wheeler 2000). Based on their interpretation of Festinger (1954), Dai and Rinn (2008) argued that “if people are certain about their abilities, there is no need to engage in social comparative information seeking” (p. 14). However, Festinger specifically hypothesized: “when an objective non-social basis for the evaluation of one's ability or opinion is readily available persons will not evaluate their opinions or abilities by comparison with others” (p. 120, Corollary IIB). We interpret this to mean that when there is a relatively objective normative basis of comparison (i.e., class- or school-average achievement, or a distribution of test results), then persons will no longer need to engage in the active social comparison strategies that have been the focus of much SCT research and highlighted by Dai and Rinn.

In addition to selecting individuals for comparison purposes, Festinger (1954) also emphasized that comparisons could be made with groups (Hypothesis VII) and that

situations could arise in which comparisons could be forced on the individual. When ability is relatively stable, Festinger proposed—apparently foreshadowing the BFLPE—that the individual would experience “failure and feelings of inadequacy with respect to this ability” (p. 137). Our interpretation of these proposals by Festinger is that when students are given accurate normative information about their performance in a particular class, social comparison information based on the performance of a specific target person should be less useful and, thus, have less influence on self-evaluations. The rationale for this interpretation, however, rests heavily on the assumption that normative comparisons typically provide more useful information in ascertaining an accurate self-appraisal, whereas it is clear that social comparison information can also be used in relation to self-serving strategies designed to protect one’s self-concept. However, as emphasized by Buckingham and Alicke (2002; also see discussion by Diener and Fujita 1997), SCT researchers have yet to clarify the relative importance of specific and generalized comparisons in evaluating one’s competencies. In summary, the role of a generalized other that is central to the BFLPE and the early development of SCT has not been such an important focus of current SCT research.

Relevance of SCT to the BFLPE

SCT theory, methodology, and empirical findings provide a heuristic basis for extending BFLPE research (and vice-versa). Nevertheless, a critical question is: How can constructs found to be important in SCT be integrated into BFLPE studies and what effects do they have? In particular, experimental manipulations, student characteristics, or group characteristics that influence the selection of the target comparisons that students choose in SCT, or the effects of this choice process, might or might not moderate the effect of school-average ability in BFLPE studies. This should be an empirical question. Although this direction for further research is implicit in the Dai and Rinn review, they sometimes imply that variables that influence individual choice of targets must necessarily influence the size of the BFLPE, or that results based on the SCT paradigm contradict the BFLPE (see ‘Evidence Constraining the BFLPE’ section of Dai and Rinn’s review), without actually testing whether SCT results even generalize to BFLPE studies or providing empirical evidence in support of their speculations.

Diener and Fujita (1997, p. 350) reviewed BFLPE research in relation to the broader SCT literature and concluded that Marsh’s BFLPE provided the clearest support for predictions based on SCT in an imposed social comparison paradigm. The reason for this, they surmised, was that the frame of reference, based on classmates within the same school, is more clearly defined in BFLPE research than in most other research settings. The importance of the school setting is that the relevance of the social comparisons in school settings is much more ecologically valid than manipulations in typical social psychology experiments involving introductory psychology students in contrived settings. Indeed, they argue (also see Marsh and Craven 2002) that except for opting out altogether, it is difficult for students to avoid the relevance of achievement as a reference point within a school setting or the social comparisons provided by the academic accomplishments of their classmates.

In BFLPE research, the implicit comparison target is posited to be a generalized other and there is a very consistent pattern of contrast effects—the negative effect of school- or class-average achievement on ASC. Even when there is evidence for assimilation effects (e.g., the pride associated with attending an academically selective high school), the net effect of attending a high-ability high school on ASC is negative (Marsh *et al.* 2000).

Furthermore, support for any assimilation effect at all—even one that is overshadowed by a counter-balancing contrast effect as in Marsh *et al.* (2000)—has been elusive (e.g., Lüdtke *et al.* 2005; Trautwein *et al.* 2006, 2008).

Similarly, based on their review of SCT research, Buckingham and Alicke (2002) noted that whereas factors such as task relevance, similarity to the comparison target, cognitive load, and perceived control may be relevant, “people generally evaluate themselves more positively when the comparison information reflects favorably (i.e., following downward comparisons) rather than unfavorably (i.e., following upward comparisons) on their characteristics and abilities, especially when they receive direct feedback regarding their own and others’ behavioral or performance outcomes” (p. 1117). However, unlike BFLPE research in which there is consistent support for a contrast effect, the theoretical predictions and empirical results are not so clear in SCT studies. In particular, when participants are able to choose a target person with whom to compare, upward comparisons sometimes result in assimilation rather than contrast, leading Major *et al.* (1991; see also Diener and Fujita 1997; Seaton *et al.* 2008c; Suls and Wheeler 2000) to describe social comparisons as a “double-edged sword” (p. 238).

An important focus of much of this SCT research has been the strategies that individuals use to select comparison targets (e.g., upward and downward comparison strategies) to maximize competing needs. Thus, upward evaluations might provide a basis of identification with more accomplished target persons even though such target persons are likely to provide a more demanding basis of comparison for self-evaluations leading to feelings of inferiority than would downward comparisons. Nevertheless, when asked to choose target persons with whom to compare themselves, SCT research shows that participants typically choose targets who are similar or slightly better than themselves (i.e., upward rather than downward; see Blanton *et al.* 1999; Huguet *et al.* 2001; Suls and Wheeler 2000).

Whereas the emphasis on generalized others in BFLPE studies may be a reasonable assumption within an imposed social comparison paradigm in educational settings (Diener and Fujita 1997), more research is needed to test the generalizability over different sources of social comparison information such as that provided by target comparison persons chosen by students in free-choice situations that has been the basis of SCT research. Furthermore, the uses of generalized and specific comparison targets are not mutually exclusive. Individuals might simultaneously evaluate their performances in relation to both the performances of specific target individuals selected in ways that have been considered in SCT research and in relation to some generalized performance based on a group-average performance, as posited in BFLPE research. In their review, Dai and Rinn (2008) put particular emphasis on two SCT studies (and research leading to these studies) that were conducted in educational settings (Blanton *et al.* 1999; Huguet *et al.* 2001), as apparently challenging the generalizability of the BFLPE. This is particularly relevant as these two studies have also been instrumental in our recent research in which we have begun to integrate the SCT and BFLPE paradigms.

Integration of BFLPE and SCT paradigms

Our research in this area began with an intriguing collaboration between some of the world’s leading SCT researchers. In preparing material for a monograph chapter on SCT (Wheeler and Suls 2004), Wheeler and Suls noted what appeared to be incompatible results based on the BFLPE and recent social psychological research, and so challenged Marsh to explain the apparent failure of BFLPE predictions in these studies (J. Suls, personal

communication, September 11, 2003). In particular, they noted that two studies (Blanton *et al.* 1999; Huguet *et al.* 2001) offered direct evidence that upward comparisons resulted in assimilation rather than the contrast effect predicted by the BFLPE. In these social comparison studies, a student's performance in a variety of academic domains was more likely to improve if they reported that they compared their exam grades with other students in their classroom who performed better than themselves (participants listed on a questionnaire their usual comparison-target in each of seven courses). Marsh was asked to reconcile the results of these studies with findings from his BFLPE research program. This challenge is closely related to concerns expressed by Dai and Rinn (2008), in relation to these same studies.

In response to this challenge, Marsh (personal communication, November 12, 2003) noted that the Blanton *et al.* (1999) and Huguet *et al.* (2001) studies did not specifically evaluate ASC and did not include a measure of class- or school-average achievement. Thus, neither study provided a test of the BFLPE nor how it related to social comparison processes that were evaluated in these two studies. Noting that class-average differences in school grades had been scaled away by standardizing grades or centering the effects separately within each class, Marsh suggested that a BFLPE might be evident for self-evaluations (which were the closest approximation to ASC that was available in these studies) if a suitable measure of class-average achievement were available, and proposed tests of the BFLPE in reanalyses of these two studies if these conditions could be established. Importantly, Marsh emphasized that these were not criticisms of the original studies in that they were not intended to test the BFLPE, but that it was also not appropriate to argue that the findings contradicted other BFLPE findings—in contrast to apparent interpretations by Dai and Rinn (2008).

In responding to Wheeler and Suls' (2004) challenge, Marsh also sent a copy of his response to authors of both the Blanton *et al.* (1999) and Huguet *et al.* (2001) studies. Independently, each research team contacted Marsh and suggested ways in which his proposed reanalyses might be undertaken. Huguet *et al.* (personal communication, April 29, 2004) suggested that all parties work together as a collaborative team to investigate the links between the BFLPE and social comparison choices. At about this time, Marjorie Seaton—who had completed her undergraduate Honour's thesis with Ladd Wheeler on social comparison theory—enrolled to do a Ph.D. under the supervision of Herb Marsh. Based on a model of collaborative synergy involving all 11 of the players in this scenario, we reanalyzed the data from both these studies in relation to predictions from the BFLPE proposed by Marsh in his original response to Wheeler and Suls. After several years of effort and large doses of good-will by all those involved, this collaborative effort eventually resulted in a publication (Seaton *et al.* 2008c) acceptable to all parties, provided apparently the first test of whether or not the BFLPE and SCT are compatible, offered an initial foundation for the integration of the BFLPE and SCT, and constituted one component of Seaton's (2007) recently completed Ph.D. thesis. Although neither of the original studies had used multilevel modeling, all the authors agreed that this was the most appropriate statistical technique to use in the reanalysis of these two studies.

Further analysis of the data of Blanton et al. (1999). Participants in the Blanton *et al.* (1999) study were 876 students from 33 ninth grade classes (the first year of high school) across four Dutch schools. Because no standardized achievement measure was available, performance was determined on the basis of school grades assessed at three points during the academic year (for further detail see Blanton *et al.* 1999; Seaton *et al.* 2008c). Comparison-level choice was measured at Time 2 by asking students to nominate the

classmate with whom they preferred to compare their grades, separately for each of seven academic subjects. For each subject, the comparison student's grade at Time 2 was then used to ascertain the comparison direction in each subject. The main dependent measure was a self-evaluation measure in which students rated their performance compared to their classmates in the seven academic subjects. In the Dutch schools, as is typically the case elsewhere (Marsh 1987), teachers tend to grade-on-a-curve such that there is not much variation between classes in terms of the average grade assigned. However, for purposes of evaluating the BFLPE, it was critical that there was a class-average measure of achievement that reflected the differing ability levels of the classes. Fortunately, each of the classes had been streamed on the basis of prior ability, and this information allowed us to scale the classes in terms of class-average ability (see Seaton *et al.* 2008c, for further information).

The Seaton *et al.* (2008c) reanalyses provided clear support for the BFLPE. For all seven school subjects, the effect of T1 grade (individual achievement) on self-evaluation was positive and statistically significant, varying from 0.32 to 0.78. The negative effect of class-average ability was significantly negative for all seven academic subjects, varying from -0.34 to -0.62 . Thus an average-ability student in a class in which the class-average mean grade was 1 SD above the mean grade of all students (in the metric of individual students), had a self-evaluation that was between -0.34 and -0.62 SDs (depending on the school subject) below the average self-evaluation across the entire sample. This re-analysis also showed that the BFLPE generalized well across ability levels, as the interactions between class-average achievement and the student's own grade were not statistically significant for any of the seven school subjects. We also juxtaposed this BFLPE with the effect of the comparison person's grade. However, the main effect of the comparison person's grade, and the interaction between class-average achievement and comparison person's grade were both statistically non-significant for all seven school subjects. Although Dai and Rinn (2008) suggest that the BFLPE is likely to be moderated by student's comparison level choice, these results show no support at all for this suggestion. Hence the BFLPE was not moderated by the choice of target student that is the focus of SCT research and there was no effect of comparison student choice on self-evaluations after controlling for the class-average ability.

Further analysis of the data of Huguet et al. (2001). Including additional students not considered in the original study of Huguet *et al.* (2001), participants were 1,156 students from 51 classes across 12 French high schools (mean age of 13.5 years). Materials were similar to Blanton *et al.* (1999) except that students only had three school subjects in common. Grades based on a 20-point scale were obtained from school reports, and were used to determine performance and comparison direction. Importantly, school grades in the French system are specifically designed to be comparable across school subjects, classes and schools (i.e., to counteract the typical grading-on-a-curve effect). For this reason and because there was no other basis of scaling the class-average achievement values for the different classes (e.g., the classes were not tracked in relation to student ability), class-average grades were used as a basis for evaluating the BFLPE in this study. However, relative to class-average differences on standardized achievement tests typically used in BFLPE studies, these class-average grades may be conservative in terms of testing the BFLPE because of a potential grading-on-a-curve effect and thus underestimate the size of the BFLPE.

As with the Dutch study (Blanton *et al.* 1999), the Seaton *et al.* (2008c) reanalyses of this French (Huguet *et al.* 2001) data provided support for the BFLPE. The effect of individual achievement was significantly positive in all three school subjects (standardized

path coefficients of 0.64 to 0.74). The negative effect of class-average ability (the BFLPE) varied from -0.15 to -0.42 , and was statistically significant for two of the three school subjects (French and Math, but not History/Geography). There were several other ways in which the Dutch and French results differed. In the analysis of the French data, the size of the BFLPE was moderated by individual ability for mathematics, but the size of this effect was small and this interaction was not significant for the other two subjects. Also, the comparison person's grade had a small positive effect on self-evaluations that varied from 0.04 to 0.12. This effect was statistically significant for two of the three school subjects, but was not statistically significant for History/Geography. Students who chose more able comparison students had higher self-perceptions of their academic ability. However, again, the comparison choice-by-class-average interaction was not statistically significant for any of the school subjects.

Juxtaposition between generalized and specific others in the BFLPE. Marsh *et al.* (2008) also sought to integrate BFLPE and SCT, comparing results based on a generalized other (operationalized as class-average achievement, as in BFLPE studies), a specific other (operationalized as direction of comparison with a freely chosen target person, as in SCT studies), and the combined effects of both these sources of social comparison information. They hypothesized that both sources of social comparison information (high class-average achievement and upward comparisons) would have negative effects on mathematical self-beliefs. However, they noted that the basis of prediction for the BFLPE was much stronger than those based on the chosen target persons. After completing a mathematics test, students nominated the student whose test booklet they would like to see and responded to the item: "Is this student in mathematics (a) better than you? (b) not as good as you? (c) similar achievement level?" The main dependent measure was a mathematics self-belief construct similar to math self-concept—mathematics agency (sample item: "When it comes to math, I'm pretty smart").

Preliminary results provided clear support for the typical BFLPE—a substantial positive effect of individual student achievement and a substantial negative effect of class-average achievement. The authors then tested the same model with the upward comparison variable (choosing an individual target of comparison who is more able) instead of class-average achievement. Results based on this alternative source of social comparison information gave a similar pattern of results. In particular, the effect of individual student achievement was positive, whereas the effect of selecting a more able student as the target of comparison (upward comparison) was negative. Finally, the authors evaluated the combined effects of both sources of social comparison information—school-average achievement (generalized other) and upward comparisons (specific other). Although the negative effects of each of these sources of social comparison information was diminished somewhat—compared to models in which each was considered separately—the negative effects of both class-average achievement and upward social comparison were significant from a statistical perspective and substantively meaningful. In summary, this set of models was consistent with a priori predictions in that the BFLPE was replicated for both class-average achievement (the typical basis of social comparison information based on a generalized other in BFLPE studies) and upward comparison (a typical basis of social comparison information based on comparison with a selected target person in SCT studies). Furthermore, when both of these sources of social comparison information were considered simultaneously in a single model, each made substantial, unique contributions.

It is important to note that each of these sources of social comparison information—the generalized other and the direction of comparison with an individual target person—

provides a unique, independent effect that cannot be explained by the other. In this sense, the single individual classmate selected by an individual student as a basis of comparison is more than just a “noisy” reflection of the class-average as a basis of comparison (i.e., a “class-average” based on a response of one randomly selected student that would obviously have considerable random error compared to the class-average based on all students within the class). Although consistent with a priori predictions based on the BFLPE, the results have important implications for both BFLPE research as well as the broader SCT literature. Importantly, the uses of generalized and specific others are not mutually exclusive alternatives. Individuals might simultaneously evaluate their performances in relation to both the performances of specific target individuals selected in ways that have been considered in SCT research and in relation to some generalized other performance based on an average performance, as posited in BFLPE research. Hence, there is a need for more research to juxtapose different operationalizations used in SCT and BFLPE research. In particular, BFLPE studies should evaluate micro-level social comparison strategies used by individual students in their selection of classmates as comparison targets, whereas SCT research should incorporate macro-level social comparison strategies based on class-average information (or some alternative representation of the “generalized other” in different settings) as well as micro-level strategies that have been the focus of this research. In relation to both operationalizations there seems to be an important role for mixed-methods research in which the largely quantitative approach used in this research is supplemented with qualitative research to more fully explicate these alternative social comparison processes. Thus, for example, it would be useful to ask students to discuss the role of social comparison in the way they form their self-concepts, upward and downward comparison strategies that they use to protect their self-concepts, the juxtaposition between normative bases of comparison based on a whole class or school and comparisons based on specifically selected individual classmates, and their perceptions of ability levels of students they chose as comparison targets. Similarly, whereas most research has focused on academic achievement (test scores and school grades), it might be interesting to examine students’ perceptions of how the comparison target student performs in other salient academic activities (e.g., board work, classroom discussion, group work, presentation of work to classmates, non-test based measures of performance, helping other students).

Summary: Alternative sources of social comparison information

Although based on very different methodologies, each of the three studies provided clear support for the BFLPE. Furthermore, the studies were consistent in showing that the BFLPE was not moderated by the direction of comparison when students were asked to choose a target person. There was, however, an important difference between the studies. In the Blanton *et al.* (1999) study, there was no effect of comparison person’s grade on self-evaluations for any of the seven school subjects. In the study of Huguet *et al.* (2001), there was a small positive effect of the comparison person’s grade on self-evaluations (an assimilation effect). However, for the Marsh *et al.* (2008) study, choosing a comparison target that was more able had a strong negative effect on math self-beliefs. Although the many differences between the studies make interpretations about the basis of these differences highly speculative, there is one important difference that warrants further consideration. In particular, in both the French and Dutch studies, the direction of comparison was inferred on the basis of differences in school grades for the target student and the comparison student, whereas in the German study the difference was based on the student’s perception of the difference between their accomplishments and those of the target

person. In the French and Dutch studies it is not known whether the target student's perception actually agreed with the differences in grades. However, given the typical optimistic bias in self-perceptions, it is likely that students overestimate their own ability relative to that of other target comparison students (also see Seaton *et al.* 2008c). In the German study, the authors did not actually know whether student self-perceptions about differences between their own ability and the ability of target comparison students were accurate in relation to objective measures of achievement. However, student self-perceptions—whether accurate or not—must be more important in determining their self-concepts than inferences about student self-perceptions based on objective measures that students do not actually see. Clearly, there is need for further research—which we are currently pursuing—that more fully explores this distinction between subjective (based on self-perceptions) and objective (based on test scores or grades) indicators of the direction of comparison.

More generally, the recent research summarized in this section is important in bringing together these two theoretical perspectives within the same study and provides important directions for further research that have not been fully explicated thus far in either BFLPE or SCT research paradigms, but are under active consideration in our ongoing research program. Clearly, important directions for further research are questions about the comparison and integration processes actually used in forming ASCs in relation to different frames of reference. In particular, it is important to explore further—perhaps using qualitative as well as quantitative methodologies—how information about the ability level of the target comparison person and the rest of the students in the class are integrated into the formation of ASC. We suspect, for example, that students with higher ASCs are more likely to select more able target comparison students with whom to compare and that part of this effect may reflect actual differences in achievement that are not captured by relatively crude measures of achievement sometimes used in this research. Alternatively, choosing a more able target of comparison may result in identification with the more able target that leads to a higher ASC. Furthermore, these possibilities are not mutually exclusive in that ASC and selection of comparison targets may be reciprocally related such that each is a cause and an effect of the other. Also, it is unclear whether individual students within the same class differ systematically in terms of how much they rely on performances of all other students within their class (a generalized other as implied in the forced comparison paradigm) and the performance of a particular target comparison person (a specific comparison person as emphasized in much SCT research) in forming their ASCs. Whereas Dai and Rinn (2008) were critical of our research in not pursuing this juxtaposition of the BFLPE and SCT paradigms, they offered no clear suggestions for directions for this research and seemed to confuse the issue by conflating research based on the BFLPE and SCT paradigms. We agree that this is an important direction for further research, but note that this has been a particularly active area of our current research which provides a solid basis for our ongoing research in this area.

Effects of Ability Tracking, Achievement Grouping, and Gifted Education Programs

In their discussion of gifted-education, Dai and Rinn (2008) note that: “Related to this issue are the effects of ability (homogeneous) grouping on self-concept as compared with that of heterogeneous grouping. Findings seem mixed in that regard (see Kulik and Kulik 1991, 1997)” (p. 12). However, their summary of research in this area is not entirely accurate. Indeed, there is a large literature on the effects of tracking, ability grouping, contextual effects, and compositional effects on diverse learning outcomes. Particularly tracking and

ability grouping are often evaluated from the perspective of contextual effects—the effects of school- or class-average achievement after controlling for the effects of individual student achievement and other individual characteristics. The main focus of this research has been on the implications of ability grouping for academic achievement. Although distinct from the effects of such ability grouping on ASC that is the focus of the BFLPE, this research on academic achievement is relevant. A comprehensive review of this research is beyond the scope of this article; however, Hattie (2002) conducted a meta-analysis that incorporated data from all existing meta-analyses of this research, providing a comprehensive summary of this research. He concluded that tracking had almost no effect on academic achievement; average effect size=0.05 (se=0.03, $n=261$ studies, 784 effects). Although there was some evidence that tracking benefited the most advantaged students in terms of academic achievement, the effect size was small (0.08), whereas the effect size was close to zero for low-tracked students. Particularly relevant to the Dai and Rinn (2008) review, Hattie emphasized that it is important to separate gifted programs from high-ability tracks when evaluating the effects of tracking. Hence, when the effect of special gifted programs was excluded, Hattie reported that the average effect size for high ability tracks was reduced to 0.02. Hattie argued that positive effects of gifted programs are due to changes in the curriculum and quality of education rather than to ability tracking per se. Many of the features of gifted programs reflect good educational practices that would likely benefit students in homogeneous classes as well.

The Dai and Rinn (2008) and Dai (2004) critiques of the BFLPE in relation to gifted education programs—a major focus of both reviews—suffer a serious flaw that was a critical issue in the Hattie (2002) review. In putting forth their argument Dai and Rinn stated that: “One can argue that gifted education provides an ideal test bed for the BFLPE theory” (p. 11) and that participating in a self-contained or short-term gifted program “gets close to the essence of the metaphor of a big fish in a little pond suddenly turned median or small when thrown into a big pond with many big or bigger fish” (p. 11). Whereas several of the studies of gifted education programs that they reviewed did show a decline in ASC consistent with BFLPE predictions (Marsh *et al.* 1995; Zeidner and Schleyer 1998), others showed no decline or declines that returned to base-line levels when students in short-term gifted programs returned to regular classes. Based on these “mixed” findings, the authors concluded that findings from gifted education research were not entirely consistent with BFLPE predictions and that consequences of participation were more complex than suggested by BFLPE theory. The flaw in this argument, as emphasized by Hattie (2002), is confounding the negative effects of ability grouping per se (the focus of the BFLPE) with the many other components that are likely to be incorporated into gifted education programs (e.g., different curriculums; more dedicated, highly trained teachers; better resources; enrichment experiences) that might be expected to have positive effects on ASC.

The fundamental flaw in the logic proposed by the Dai and Rinn (2008) is their implicit assumption that support for the BFLPE theory necessitates that individual student achievement and school- or class-average ability are the only variables that influence ASC. Although BFLPE theory does predict the negative effects of school- or class-average ability, it clearly does not assume that this is the only influence on ASC or that these negative effects cannot be mediated, counter-balanced, or moderated by other influences. In this respect, gifted education programs typically confound the potentially positive effects of many aspects of gifted education programs with the negative effects of ability grouping that are the focus of the BFLPE. Indeed, the fact that the preponderance of results in the Dai and Rinn (2008) critique show that the effects of gifted education programs on ASC are

negative—despite the many aspects of such programs that might be expected to enhance ASC—seems to provide strong support for predictions based on the BFLPE. However, these potentially positive and negative effects are not easily unconfounded in gifted education programs, suggesting—in contrast to suggestions by Dai and Rinn—that these studies do not provide an ideal test of the BFLPE. In an interesting variation on the typical BFLPE study that addresses this issue in part, Preckel *et al.* (2008) evaluated the BFLPE considering only gifted-education classes, thus controlling many of the potentially confounding factors in BFLPE with gifted education factors. Noting there is considerable variation in the class-average achievement levels even within gifted education classes, they found that there was a substantial negative effect of class-average ability. Hence, even in a sample limited to gifted education classes, there is support for the BFLPE.

Dai and Rinn (2008), in the same section of their article (“Applications of the BFLPE Theory to Attending Gifted Programs” p. 11–14), also reviewed results from meta-analyses by Kulik and Kulik (1991, 1997; also see Kulik and Kulik 1982) on the effects of ability grouping. The conclusion of these meta-analyses was that ability-grouped students (compared to non-ability-grouped students) did not differ systematically in terms of self-concept. Dai and Rinn argued that results of these meta-analyses were inconsistent with BFLPE predictions—although noting limitations of these results that sometimes were based on non-academic components of self-concept or self-esteem. However, in response to the original Kulik and Kulik (1982) meta-analysis, Marsh (1984b) pointed out that their meta-analysis confounded negative effects for high-track students with positive effects for low-track students—both of which are consistent with BFLPE predictions. In a subsequent reanalysis of their results, Kulik (1985) reported that Marsh’s predictions based on the BFLPE were supported when the effects of high-track and low-track students were considered separately. This same pattern of results—counterbalancing positive effects in low-tracks and negative effects in high track—was presented in greater detail in the more comprehensive Hattie (2002; also see Trautwein *et al.* 2006) review of meta-analyses of ability grouping research that incorporated the meta-analyses by Kulik and Kulik. In summary, whereas we agree with Dai and Rinn about the need to more carefully distinguish between academic and non-academic self-concept, a more critical evaluation of these meta-analysis results is consistent with the BFLPE and not consistent with apparent interpretations offered by Dai and Rinn.

Generalizability of the BFLPE

Moderation effects

Dai and Rinn (2008) claimed that the theoretical model underlying the BFLPE assumes that the negative effect of school-average achievement is invariant and thus is not sufficiently complex to take into account other influences that might moderate the size of the negative effects of school-average ability. However, contrary to claims by Dai and Rinn, the actual theoretical model underlying the BFLPE does not make this assumption. Dai and Rinn (2008) imply that BFLPE research treats the effect of school-average achievement as invariant, ignoring individual differences, cultural differences, and contextual features (other than school-average ability) suggested to be important in other theoretical frameworks such as SCT and motivation theory. Based on this erroneous implication, they go on to conclude “most of the BFLPE studies are indiscriminate of contextual features other

than school-average ability or achievement, which is typically the only basis for estimating the BFLPE” (p. 27). However, they also argue that (p. 27):

In general, the research strategy of the BFLPE program is to show generality and ubiquity of the BFLPE over gender, ability levels, and cultures (Marsh and Hau 2003; Marsh *et al.* 2007), rather than finding out details of how it works psychologically (i.e., addressing the issue of internal validity), as evidenced by their preference for large-scale data sets and statistical manipulation to tease out the effects (e.g., Marsh 1987, 1994; Marsh and Hau 2003; Marsh *et al.* 2007). (p. 27)

Hence, the nature of Dai and Rinn’s arguments appear to be internally inconsistent. On the one hand they suggest that BFLPE research assumes the effect of school-average ability to be invariant and ignores potential moderating variables. On the other hand they acknowledge that BFLPE studies have systematically evaluated the extent to which the effect of school-average ability varies with other individual difference and contextual variables. The underlying criticism seems to be not that BFLPE studies have ignored the potential moderating effects, but that the sizes of these interactions have been consistently small (or nonsignificant) and not even consistent across different studies (see “Summary” section of Dai & Rinn’s paper). Indeed, it seems strange to argue that the high level of generalizability and robustness of the basic BFLPE predictions should be seen as a limitation in the theory. Typically, generalizability is seen as a strength rather than a weakness.

Importantly, claims by Dai and Rinn (2008) that such interactions have been ignored and are inconsistent with BFLPE theory are inaccurate. There is nothing inherent in the theoretical model of the BFLPE that argues that the effect cannot be moderated, nor is there any theoretical basis for arguing that the effects are necessarily invariant. Quite the contrary, many BFLPE studies are based on the premise that the BFLPE does interact with individual or contextual level variables like those discussed by Dai and Rinn. Thus, for example, in suggestions particularly relevant to gifted education settings that are the focus of the Dai and Rinn critique, Marsh (1993, 2007; Marsh and Craven 1997, 2002; Seaton *et al.* 2008a) argued that the BFLPE should vary systematically with some individual characteristics as well as strategies that might partially counter the BFLPE. They proposed that the BFLPE might be moderated by motivational orientations (e.g., competitive vs. mastery) and climates, use of individualized assessment tasks, avoiding competitive climates that encourage social comparison, feedback in relation to criterion reference standards and personal improvement over time, and reinforcing identification with other participants to enhance reflected glory effects. Nevertheless, although nearly all BFLPE studies have evaluated the extent to which the size of the negative effect of school-average ability interacts with other variables, the results suggest that such interactions are small (or nonsignificant) and inconsistent across studies; those that have been found influence the size of the BFLPE but not its direction.

Dai and Rinn (2008) were critical of BFLPE studies for not seeking process variables that moderate the BFLPE and for not extending research to include systematic classroom observation, but failed to acknowledge BFLPE studies that did. For example, Dai and Rinn cited the Lüdtkke *et al.* (2005) study, but failed to acknowledge that it was based, in part, on classroom observation and was specifically designed to test the hypothesis that an important teaching style (individualized teacher frame of reference, TFR) moderated the BFLPE. Teachers with an individualized TFR emphasize improvement in relation to prior achievement, effort, and learning. TFR was independently assessed by student ratings of their teacher and ratings by two trained observers. Lüdtkke *et al.* hypothesized that this teacher level variable would have a positive effect on ASC and moderate the BFLPE such

that students in classes where the teacher had a high TFR would have smaller BFLPEs. Based on 2,150 German students from 112 classes, multilevel analyses replicated the BFLPE (the negative effect of class-average achievement) and the positive effects of TFR on ASC. However, TFR did not affect the size of the BFLPE, and this result was consistent across both student and observer ratings of TFR. In summary, this carefully conducted study based on actual classroom observation as well as student ratings failed to support the hypothesis that the BFLPE would be diminished by more appropriate teaching styles designed to counteract the BFLPE. Importantly, however, the results do support the positive effects of an individualized TFR in relation to enhancing ASC.

This issue of moderated effects in relation to the BFLPE was one focus of the recently completed Seaton (2007; also see Seaton *et al.* 2008b) Ph.D. thesis that used the PISA2003 data (nearly a quarter million students from 41 countries). Across the 41 countries and for each country considered separately, the results largely replicated the earlier Marsh and Hau (2003) study based on the 26 countries in PISA 2000. Extending this research to include 41 countries and a larger sample of non-Western countries, Seaton showed that the size of the BFLPE generalized across non-Western countries and collectivist cultures, as well as Western countries and more individualistic cultures (also see Seaton 2007). Seaton also extended the earlier research by including a variety of moderating variables that might be expected to moderate the size of the BFLPE. A number of moderators were statistically significant (due the extremely large sample size) but small in size, students suffered slightly less from the BFLPE if they: (a) used elaboration techniques; (b) were more extrinsically motivated; (c) were more intrinsically motivated; (d) felt a sense of academic self-efficacy; (e) had more positive attitudes to school; (f) felt a connection to the school; or (g) came from high SES families. BFLPEs were somewhat larger for students who used memorization strategies, or who preferred cooperative learning environments. Whereas all of these effects were very small and would probably have been non-significant in even moderately large samples, one effect was sufficiently large to be substantively important: highly anxious students experienced larger BFLPEs. Even here, however, the direction of the BFLPE was consistent across levels of anxiety; even low-anxious students showed BFLPEs although smaller in size than those found with high-anxious students. Furthermore, the interpretations are complicated by the fact that other research (Zeidner and Schleyer 1998) suggest that students in academically selective classes have systematically higher levels of anxiety, so that more research is needed to disentangle the effects of school- and class-average achievement on academic self-concept and text anxiety.

The most extensive research has been done on interactions between school-average ability and individual ability—evaluating whether the size of the BFLPE varies with the academic ability of individual students. Marsh (1984a, 1987, 1991; Marsh and Craven 1997; Marsh *et al.* 1995; Marsh and Rowe 1996) argued that attending high-ability schools should lead to reduced ASCs for students of all achievement levels based on several different theoretical perspectives. For a large, nationally representative (US) database, Marsh and Rowe (1996) found that the BFLPE was clearly evident for students of all achievement levels and that the size of the BFLPE varied only slightly with individual student achievement. In two studies demonstrating BFLPEs in students attending gifted-and-talented programs, Marsh *et al.* (1995) found no significant interaction between the size of the BFLPE and achievement level of individual students. In their cross-cultural study of the BFLPE in 26 countries, Marsh and Hau (2003) also found that the BFLPE did not vary with individual achievement levels. In their review of BFLPE research, Marsh and Craven (2002) concluded that there is little evidence that the size of the BFLPE varies systematically with individual student ability levels. Hence, the BFLPE generalizes well over different student ability levels.

Relations with other outcomes

The clear support for apparently paradoxical predictions based on the BFLPE is exciting for self-concept researchers, but what are the policy implications of these findings and how do the results generalize to other outcomes? Marsh (1991) considered the influence of school-average achievement on a much wider array of outcomes in the large, nationally representative, longitudinal High School and Beyond study of US high school students surveyed in Year 10, Year 12, and again 2 years after graduation from high school. The High School and Beyond outcomes were specifically designed to include most of the important outcomes of education. After controlling for background and initial achievement, the effects of school-average achievement were negative for almost all of the Year 10, Year 12, and post-secondary outcomes: 15 of the 17 effects were significantly negative and two were non-significant. School-average achievement most negatively affected ASC (the BFLPE) and educational aspirations, but school-average achievement also negatively affected general self-concept, advanced coursework selection, school grades, academic effort, standardized test scores, occupational aspirations, and subsequent college attendance. The negative effects for educational aspirations were clearly evident 2 years after graduation from high school. Controlling for the negative effects of school-average achievement on ASC substantially reduced the size of negative effects on other outcomes, consistent with the proposal that these negative effects of school-average ability were substantially mediated by ASC. These results suggest that the negative effects of attending high ability schools extend well beyond those for ASC that has been the focus of BFLPE studies.

Other recent research shows that the BFLPEs have long-lasting effects on other variables in addition to the negative effects on ASC. Thus, for example, Marsh and O'Mara (2008) showed that school-average ability early in high school not only had negative effects on ASC (the BFLPE), school grades, and educational and occupational aspirations during high school, but continued to have negative effects up to 5 years after high school graduation. In physical education settings, Chanal *et al.* (2005) showed that the BFLPE generalized to gymnastics self-concept in a gymnastics training program, whereas Trautwein *et al.* (2008a) found that class-average physical ability not only had a negative effects physical self-concept but also had a negative effect on longterm physical activity levels.

Trautwein *et al.* (2006) extended BFLPE research to consider academic interest (intrinsic value, personal importance, and attainment value), noting that research into interest had largely ignored frame-of-reference effects. Juxtaposing ASC and interest, they found—consistent with BFLPE predictions—that both constructs were positively influenced by individual student achievement and negatively influenced by school-average achievement. Next they asked what was the process underlying these results. Consistent with predictions based on expectancy-value theory (Eccles 1983) and the Marsh *et al.* (2005) longitudinal study of the causal ordering of ASC and interest, they found that ASC almost completely mediated the BFLPEs on interest.

In summary, there is consistent support for the negative effects of school-average achievement on ASC—the BFLPE. Although the BFLPE refers specifically to effects on ASC, there is a growing body of research suggesting that school-average ability also has negative effects on a variety of other variables such as the studies reviewed here. However, this research tends to be idiosyncratic; there is a need to develop a theoretical framework and conduct systematic research about when these effects are likely to be negative and how they relate to the BFLPE. Particularly useful would be more longitudinal studies that

juxtapose the effects of school-average achievement on ASC and other variables, but also attempt to test the causal ordering of these effects. Clearly there exists some research suggesting that many of the long-term effects of school-average ability on other constructs are substantially mediated by ASC, attesting to the importance to the BFLPE and the potency of ASC as an important outcome variable in education that facilitates the attainment of many other desirable outcomes.

BFLPE stability over time

Dai and Rinn (2008; Dai 2004) suggested that the BFLPE might be a short-term, ephemeral effect, arguing that the lack of apparent long-term negative self-related or motivation-related consequences “challenge the external or ecological validity of the BFLPE model” (p. 14). However, there is good empirical evidence to counter this claim. Indeed, the size of the BFLPE typically remains stable or even increases in size over time for students who remained in the same school setting (Marsh and Hau 2003; Marsh 2005a; Marsh and Craven 2002; Marsh *et al.* 2007). For example, in the large US High School and Beyond Study, Marsh (1991) demonstrated that there were new BFLPEs experienced in the final year of high school beyond those already experienced earlier in high school.

The Marsh *et al.* (2001) German study of the reunification of East and West German school systems was particularly important in demonstrating the temporal evolution of the BFLPE. They found that the size of the BFLPE increased substantially during the first year after reunification for East German students who had not previously experienced selective schools compared to West German students who had previously attended selective schools for the 2 years prior to the reunification. For East German students the BFLPE was not evident at the start of the school year, had grown larger but was still less than for West German students by the middle of the school year, and was as large as the BFLPE for West German students by the end of the school year. Hence, the onset of the BFLPE was gradual, taking at least half a school year to be evident. This time frame is particularly relevant in that several gifted education studies cited by Dai and Rinn (2008) are based on very short programs, sometimes lasting only a few weeks.

In a large Hong Kong study of students entering selective schools in Grade 7 (Marsh *et al.* 2000), there was a substantial negative effect of school-average ability in Grade 9 even after controlling for the substantial negative effects in earlier school years. In this longitudinal study there were extensive pretest standardized achievement measures available for all students prior to the start of high school that were part of the selection process used to determine the high school that students would be able to attend. Hence, school average-ability measures were based on an extensive battery of tests collected prior to the start of high school, facilitating causal interpretations of the BFLPE and demonstrating its growth over time.

The negative effect of school-average ability seems to grow more negative the longer a student remains in the same school. A more demanding challenge is to evaluate the stability of the BFLPE on ASC several years after graduation from high school, when the frame of reference based on other students from their high school is not so salient and is no longer imposed by the immediate context. Extending this work on the stability of the BFLPE over time, two recent German studies (Marsh *et al.* 2007) showed that the substantial BFLPE at the end of high school showed little or no diminution 2 years (Study 1) or 4 years (Study 2) after graduation from high school. Marsh and O'Mara (2008) took a somewhat different perspective to this issue in a longitudinal analysis of responses collected on five occasions over eight critical developmental years (grade 10 to 5 years after high school graduation). School-average-ability had negative effects on ASC (the BFLPE), school grades,

educational and occupational aspirations, and educational attainment. Previous research has typically reported short-term negative direct effects of school-average ability, but using complex structural equation models, the authors demonstrated that long-term total (direct plus indirect) negative effects of school-average ability were systematically much larger than direct effects across diverse educational outcomes, and explored how the effects of school-average ability on long-term distal outcomes were mediated through effects on more proximal variables. Applying a new, stronger methodological approach that is more broadly appropriate for longitudinal and developmental research, they showed how the size of the total BFLPE—including indirect effects—has typically been underestimated in previous longitudinal studies.

Hence, in contrast to suggestions by Dai and Rinn (2008; Dai 2004), longitudinal studies demonstrate that the BFLPE is not a short-term, ephemeral effect. These studies demonstrate that as long as students remain in the same high school and the school-average achievement is relatively stable so that the immediate frame of reference remains reasonably consistent, there is ample evidence that the BFLPE persists or even increases in size. This is not surprising, and is consistent with the rationale underpinning the imposed social comparison paradigm posited by Diener and Fujita (1997). Furthermore, the direct effects of school-average ability that are the basis of most BFLPE studies are likely to substantially underestimate the total effects, particularly in longitudinal studies with many waves of data over an extended period of time.

Methodological Implications: Current Progress and Future Directions

Dai and Rinn (2008) acknowledge research methodology strengths of the BFLPE research, but argue that “the methodology (including statistical designs and data collection methods) reveals weaknesses and flaws” (p. 27). Here we address some of their main claims, but also point to directions for future research. We begin with a brief overview of the methodological approaches that have been implemented in BFLPE studies and then address specific claims by Dai and Rinn.

Methodological approaches to the BFLPE: a substantive–methodological synergy

Complex substantive issues require sophisticated methodologies—a substantive–methodological synergy (Marsh and Hau 2007). The rapid development in quantitative methods has enabled researchers to explore previously inaccessible problems, revisit classic unresolved issues with stronger tools, and address new issues—but only if substantive research incorporates new methodological tools that are appropriate.

Historically (e.g., Marsh 1984a, b, 1987, 1991), BFLPE research was based on single level models. In the earliest application (Marsh and Parker 1984) the BFLPE was based on a single-level model based on manifest scores, using a small number of schools. By current standards, this was clearly unacceptable. In subsequent applications, Marsh (1987, 1991) again used a single-level multiple regression with manifest variables. However, the numbers of schools (88) was much larger and he used a crude estimate of a design effect to compensate for the clustered sampling. Marsh (1994) then applied a single-level SEM in which key constructs were measured with multiple indicators, the number of schools was large, and a crude design effect was used to correct standard error estimates.

Marsh and Rowe (1996) was apparently the first BFLPE study to use a true multilevel analysis in a reanalysis of the Marsh (1987) data using a true (two-level) multilevel

approach based on manifest indicators. Subsequent BFLPE studies (Marsh *et al.* 2000, 2007, 2008; Marsh and Hau 2003) have been based on multilevel models with two levels in which L2 was either school or class, depending on the design of the study. Marsh and Hau (2003; Seaton *et al.* 2008a, b) subsequently applied a three level model (level 1 = students, level 2 = schools, level 3 = countries) with OECD/PISA data to test the cross-national generalizability of the BFLPE. In a recent BFLPE study with a particularly complex factor structure (19 constructs inferred from multiple indicators measured over an 8 year period), Marsh and O'Mara (2008) implemented the “complex design” option available in the *Mplus* statistical package instead of a multilevel model. In that study, both individual student and school-average variables were based on multiple indicators, and the analyses took into account the clustered nature of the data.

BFLPE research—like most applied social science research—has focused on either SEMs or multilevel analyses, but has not fully integrated the two into a single analytic framework. In multilevel analyses that have dominated recent BFLPE research, ASC, achievement and other constructs are based on manifest indicators (e.g., scale scores) even when there are multiple indicators of each construct. An implicit, unwarranted assumption in these analyses is that these student level (L1) constructs are measured without error, resulting in underestimation of their effects and complicated implications for estimated effects of school-level variables. For such aggregations of L1 constructs, Lüdtke *et al.* (2007, 2008) showed that the unreliability of the school mean can lead to biased estimation of contextual effects, particularly when the number of observations per school is small and when the intraclass correlation of the corresponding student observations is low. They introduced a latent covariate approach that regards the unobserved school mean as a latent variable, consistent with the reflective aggregations of L1 constructs. To the extent that there is measurement or sampling error in the use of observed class-average achievement, existing BFLPE research is likely to underestimate the size of the BFLPE based on new methodological approaches that control for unreliability in aggregated L2 constructs. Although these new developments in the analysis of multilevel latent contextual models have not yet been applied to BFLPE research, such possibilities provide important directions for further research that are actively being pursued in our research program.

Methodological criticisms by Dai and Rinn (2008)

Lack of specification of contexts Dai and Rinn (2008) argued: “There is no specification of the contexts where the BFLPE is more or less likely to occur” (p. 27), that the ability to look at these effects is limited due to the use of large-scale data based, and that most BFLPE studies are “indiscriminant of contextual effects other than school-average ability” (p. 27). As already discussed in relation to “moderated effects” this claim is untrue. Numerous BFLPE studies have posited characteristics of the individual student, the teacher, the classroom, and the school that are likely to influence the size of the BFLPE and systematically evaluated these predictions. Whereas we readily acknowledge the value of classroom observations, qualitative data, and case study designs to address these issues, it is clear that contextual variables can—and typically are—included in many large-scale databases used in BFLPE studies and that sophisticated statistical analyses are needed to interrogate the interpretations of these contextual effects.

Implicit specification of social comparison Dai and Rinn (2008) claim that in BFLPE studies “social comparison is inferred” (p. 28), there is no direct evidence that students

“engage in social comparison” (p. 29) and “it is difficult to know whether the BFLPE is due to more downward comparison in less selective schools or more upward comparison in more selective schools” (p. 28). There is clear evidence from BFLPE and SCT research that students do compare their academic accomplishments with those of other students and use this as one basis for forming their self-evaluations (Blanton *et al.* 1999; Diener and Fujita 1997; Huguet *et al.* 2001; Marsh and Craven 2002; Seaton *et al.* 2008c; Suls and Wheeler 2000). Hence the suggestion that there is no evidence that students do engage in social comparison is unwarranted. The need to distinguish between upward and downward comparison processes highlighted by Dai and Rinn is highly relevant in the traditional SCT theory where participants are given considerable flexibility in choosing comparison targets, the strategies that they use, and the implications of upward and downward comparisons. This issue is less relevant to the BFLPE in which there is an implicit assumption that all students within a given context compare themselves with a normative average value representing that context. In this sense, it is the juxtaposition between the student’s own accomplishments and the normative average value that is one important determinant of ASC—not the social comparison selection processes that students use to select individual students with whom to compare themselves. Indeed, the surprising result is that that BFLPE is so robust, generalizing across a range of individual student-, class-, and school-level constructs that might be expected to influence social comparison processes. Nevertheless, we agree that there is need for more research to integrate micro-level processes that are the focus of SCT and the more macro-level processes that are the focus of BFLPE research that extends Seaton *et al.* (2008c). However, evidence so far suggests that the BFLPE is not moderated by the direction of selection in the traditional SCT paradigm.

Statistical issues and effect sizes Dai and Rinn (2008) argue that statistical analyses in BFLPE studies are not rigorous, are subject to artifact, and provide a weak basis for inferring causality. BFLPE studies—and contextual models more generally—are largely based on correlational analyses so that causal interpretations should be offered tentatively and interpreted cautiously. Here, as with all social science research, it is appropriate to hypothesize causal relations but researchers should fully interrogate support for causal hypotheses in relation to a construct validity approach (see Marsh 2007) based on multiple indicators, multiple (mixed) methods, multiple experimental designs, multiple time points, and testing the generalizability of the results across diverse settings. Whereas stronger inferences about causality are possible in longitudinal, quasi-experimental, and true experimental (with random assignment) studies, trying to “prove” causality is usually a precarious undertaking. Even in true experimental studies in applied social science research, there is typically some ambiguity as to the interpretation of what was actually manipulated and its relevance to theory and applied practice.

Fortunately, there is now a growing body of BFLPE research that addresses many of these concerns. Quasi-experimental, longitudinal studies based on matching designs as well as statistical controls show that ASC declines when students shift from mixed-ability schools to academically selective schools—over time (based on pre-post comparisons) and in relation to students matched on academic ability who continue to attend mixed-ability schools. For example, in the Marsh *et al.* (2000) Hong Kong study, school-average ability was based on a pretest battery of test scores collected prior to the start of high school so that there was no possibility that school-average ability measures were confounded with academic growth attributed to attending academically selective high schools. Extended longitudinal studies show that BFLPEs grows stronger the longer students attend selective schools and are maintained even 2 and 4 years after graduation from high school.

There is good support for the convergent and discriminant validity of the BFLPE as it is largely limited to academic components of self-concept and nearly unrelated to non-academic components of self-concept and to self-esteem. School-average math ability has a much stronger negative effect on math self-concept than verbal self-concept, whereas school-average verbal ability has a much more negative effect on verbal self-concept than math self-concept. Cross-national comparisons based on OECD-PISA data from representative samples from many countries shows that the BFLPE has good cross-national generalizability.

Whereas the “third variable” problem is always a threat to contextual studies that do not involve random assignment, Marsh *et al.* (2004) argue that this is an unlikely counter-explanation of BFLPE results in that most potential “third variables” (resources, per student expenditures, SES, teacher qualifications, enrichment experiences, etc) are positively related to school-average achievement, so that controlling for them more effectively would increase the size of the BFLPE (i.e., the negative effect of school-average achievement). In this respect, BFLPEs are conservative in relation to this concern.

Dai and Rinn (2008) argued that the relatively low effect sizes associated with the BFLPE are not compelling and suggest a “host of intervening factors moderating and mitigating the alleged negative effects of school selectivity on academic self-concepts” (p. 32). In fact, Dai and Rinn did not present any actual effect sizes based on the BFLPE and apparently confused the size of regression coefficients with effect sizes. Tymms (2004; Trautwein *et al.* 2008a) proposed the effect size for continuous level-2 predictors in multilevel models, which is comparable with Cohen’s *d* (1988), be calculated using the following formula:

$$\Delta = 2 \times B \times SD_{\text{predictor}} / \sigma_e$$

where *B* is the unstandardized regression coefficient in the multilevel model, $SD_{\text{predictor}}$ is the standard deviation of the predictor variable at the class level, and σ_e is the residual standard deviation at the student level. Applying this approach to the PISA 2003 data (Seaton 2007, p.137; also see Seaton *et al.* 2008b) in the largest, most representative test of the BFLPE to date, the effect size for the total sample was 0.49. We also note that the standardized regression coefficients reported by Dai and Rinn (2008) substantially underestimate effect sizes like those traditionally used in research reviews and meta-analyses. Hence, the effect size for the BFLPE is clearly large enough to warrant practical attention as well as being substantively and theoretically important.

Finally, Dai and Rinn (2008) suggested a variety of new and different analytic strategies and research designs that could be applied to BFLPE research. Although this type of generic criticism could be made to almost any applied area of social science research, some of their specific suggestions are inappropriate. Thus, for example, they stated: “we suggest the use of individual growth modelling as an alternative to multi-level modelling” (p. 37). Although we applaud the use of growth modeling in BFLPE research, it must be incorporated into multilevel models—not used instead of multilevel models. This caveat would also apply to the recommended application of more idiographic quantitative approaches like latent-class and latent-profile analysis. More generally, we find it ironic that this criticism about not incorporating new methodological approaches is leveled at BFLPE. Clearly, as shown by the research reviewed here, BFLPE research reflects a methodological–substantive synergy in which substantive findings are based on state-of-the-art statistical analyses and questions raised from substantive research make a contribution to methodology—and will continue to do so.

We also note that Dai and Rinn (2008) seem to imply that the typical multilevel regression model used to test the BFLPE would not be appropriate for latent growth modeling, latent class analysis, and experimental (or quasi-experimental) designs with experimental and control groups. However, this is clearly not the case. New and emerging analytic strategies allow researchers to integrate multilevel modeling with latent growth analysis; to ignore the multilevel structure of data when applying these new approaches would be a serious limitation. Furthermore, even in a true experimental design it is easy to include the experimental groups (represented by a dichotomous variable if only two groups, or contrasts if more than two groups) in a multilevel regression analysis—using extensions of well known multiple regression approaches to ANOVA. This more general approach also allows inclusion of multiple indicators for the different variables considered in the analysis, thereby controlling for measurement error in a way that cannot easily be accomplished in traditional (single-level) ANOVA and multiple regression analyses used in many studies cited by Dai and Rinn. However, if there are multiple classes in each of the experimental groups, it is still important to include class-average achievement to determine whether the experimental manipulation has any effect on ASC beyond what can be explained in terms of the BFLPE. Thus, for example, several studies have shown that the type of class (i.e., high-ability or not) has little or no effect on ASC beyond what can be explained in terms of class-average ability (e.g., Marsh *et al.* 2000; Marsh *et al.* (2001); Trautwein *et al.* 2006) thus supporting interpretations of the BFLPE and the implications of the intervention. If there are not multiple classes in each group, it is still advantageous to use a latent-variable model to analyze the results, but it is important to recognize the limitations in terms of generalizability of such a case study approach with $N=1$ class in each group. Indeed so-called experimental studies in which a small number of intact classes are randomly assigned to different treatment groups typically do not provide an adequate basis for testing experimental hypotheses.

In summary, the results of any one BFLPE study are likely to provide a limited basis of support for research hypotheses positing causal effects that must be examined in relation to a broadly conceived construct validity approach. Although fraught with philosophical and methodological conundrums—including those identified by Dai and Rinn and others identified here—many of these issues have been addressed through the accumulated research evidence from BFLPE studies. Nevertheless, we welcome the opportunity to explore further the construct validity of interpretations of the BFLPE and look forward to results of research by Dai, Rinn and colleagues that pursues some of their suggestions in more detail.

Competitive Environments and Speculations on How to Counter the BFLPE

In a highly competitive environment there are likely to be a few “winners,” a lot of “losers,” and a general decline in self-concept (Covington 2001). Hence, Marsh and Craven (2002) speculated that the BFLPE could be reduced by de-emphasizing highly competitive environments that encourage the social comparison processes: Develop assessment tasks and feedback that encourages individual students to pursue their own projects that are of particular interest to them to reduce social comparison; Provide students with feedback in relation to criterion reference standards and personal improvement over time rather than comparisons based on the performances of other students; and Emphasize to each student that she or he is a very able student and value the unique accomplishments of each individual student so that all students can feel good about themselves. Whereas such

strategies were proposed specifically to undermine the negative BFLPE in high-ability schools, Marsh and Craven noted that these strategies also reflected good teaching that should improve educational outcomes generally. Although heuristic, it is important to emphasize that there is little empirical support for the strategies offered by Marsh and Craven. Indeed, as emphasized here, most of the research has found that the BFLPE is very robust, generalizing over a range of individual student characteristics and classroom climate variables that might be expected to moderate the size of the BFLPE. However, there has been very little research in which classroom or teacher-level variables have been experimentally manipulated in true experimental or quasi-experimental studies specifically designed to counter the BFLPE.

Some indirect support for Marsh and Craven's (2002) speculations comes from a physical education intervention. Marsh and Peart (1988) constructed two different physical education programs that experimentally manipulated the type of performance feedback given to high school girls who were randomly assigned to one of two experimental groups or a no-treatment control group. Participants completed a physical fitness test and a self-concept instrument prior to, and immediately following, a 6-week intervention consisting of fourteen 35-min classes. The two experimental groups participated in aerobics training programs that differed in the nature of tasks, feedback, and motivational cues given to students. The social-comparison/competitive feedback emphasized the relative performances of different students and focused on whoever performed best for a particular exercise, whereas the improvement/cooperative feedback emphasized progress in relation to previous performances. In the social-comparison/competitive group, all the physical activities were done individually. In the improvement/cooperative group, the activities were done in pairs so that one student could not succeed without cooperation with a partner. Both experimental interventions significantly enhanced physical fitness relative to pretest scores and in comparison to the control group; there were no differences between these two experimental groups in terms of gains in fitness. The improvement feedback intervention also significantly enhanced physical self-concept, but the social-comparison intervention produced a significant decline in physical self-concept. Apparently, the social-comparison feedback forced participants to compare their own physical accomplishments with the participants who were best on each individual exercise to a much greater degree than had been the case prior to the intervention or in the control group. Even though students in the social comparison condition had substantial gains in actual fitness levels, these gains were more than offset by the much more demanding standards of comparison forced upon them in the classroom environment. Although there was no long-term follow-up, Marsh and Peart speculated that the diminished physical self-concepts in the social comparison/competitive group would undermine initiative to pursue further physical activity needed to maintain the enhanced physical fitness. Hence, this study demonstrates that the nature of feedback given to students can fundamentally affect self-concept in a way that is consistent with speculations offered by Marsh and Craven (2002) on how to counter-act the BFLPE. Clearly, classroom-based experimental interventions of this sort are an important direction for further research to better test strategies about how to counter the negative consequences of the BFLPE.

Summary

We agree with many issues raised by Dai and Rinn (2008). Indeed, we have already incorporated some of their suggestions into our ongoing research program. However, as

emphasized in this review, we feel that Dai and Rinn (2008; Dai 2004) have misconstrued some aspects of the BFLPE; confused and confounded theoretical, methodological, substantive findings based on the BFLPE and SCT; and sometimes critiqued BFLPE research based on their misinterpretations of BFLPE research rather than actual BFLPE theory and research. They argue that SCT provides findings contradictory to the BFLPE, but provide little or no empirical evidence about how—or even if—these SCT findings are relevant and generalize to the BFLPE. Importantly, our recent research on the juxtaposition between SCT and the BFLPE (Seaton 2007; Seaton *et al.* 2008c) summarized here shows that their speculations are largely inaccurate.

Dai and Rinn (2008; Dai 2004) argue for the need for further research into potential moderators and mediators of the BFLPE. We applaud pursuit of this research, but reject the claim that such evidence would necessarily undermine BFLPE theory and research. Furthermore, BFLPE studies have pursued a wide variety of potential moderators—including many proposed by Dai and Rinn and others as well—but has not found any individual student or contextual variables or processes that substantially moderate even the size of the BFLPE—and certainly not its direction. Indeed, Dai and Rinn seem to imply that this robustness is a weakness in BFLPE theory and research, whereas we interpret it to be a strength.

Dai and Rinn (2008) argue that it would be useful to apply new analytic strategies and alternative experimental designs (e.g., longitudinal, growth modeling, latent-class analysis, and idiographic research) to BFLPE research. Again we would welcome such research and have consistently applied and developed new analytical approaches in our own research to extend the state of the art of BFLPE research. However, this endorsement of the need to apply new and evolving methods does not necessarily undermine support for existing BFLPE theory and research. Rather, as we have found when we apply new, stronger analytic techniques, we suggest that pursuit of their suggestions will complement and strengthen current BFLPE research.

Dai (2004) argued that the BFLPE is a short-term ephemeral effect, and Dai and Rinn (2008) still seem to have lingering doubts about the stability of the BFLPE. However, countering any such suggestions is new and previous evidence from our longitudinal research showing that the size of the BFLPE is stable or grows larger over time. Particularly in the area of gifted education, Dai and Rinn's argument that any gifted education program that increased self-concept would undermine support for the BFLPE is fundamentally flawed unless such studies are able to disentangle the apparently negative effects of ability grouping that is the focus of the BFLPE from that potentially positive features that are likely to be incorporated into gifted education programs (e.g., different curriculum; more dedicated, highly trained teachers; better resources; enrichment experiences). Indeed, the fact that the preponderance of gifted-education results in the Dai and Rinn critique show negative effects of gifted education programs on ASC—despite the many aspects of such programs that might be expected to enhance ASC—seems to support the robustness of the BFLPE. Furthermore, Dai and Rinn misinterpreted the related results of meta-analyses of ability grouping studies as failing to support the BFLPE. However, a more careful evaluation of the pattern of results shows negative effects of high-track grouping and positive effects of low-track grouping on ASC—results that are consistent with the BFLPE as has been noted in several previous reviews of this literature.

In summary, we applaud many of the suggestions by Dai and Rinn (2008) as to how BFLPE research could be extended, as evidenced by the fact we have previously proposed similar directions for further research (e.g., Marsh and Craven 2002), have actually implemented some of them in research reviewed here, and will continue to pursue these and

other suggestions in our ongoing research program. Furthermore, we welcome the opportunity to defend our interpretations of the BFLPE from a broad construct validity perspective and discuss further research that is needed. Certainly we agree with Dai and Rinn that more research is needed. We hope that this interchange will motivate them and others to pursue this further BFLPE research, and challenge us to refine our methods as appropriate. More generally, we appreciate the vigorous and vibrant debate that our research has stimulated and firmly believe that such debate will broaden the scope and strengthen BFLPE research.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Alwin, D. F., & Otto, L. B. (1977). High school context effects on aspirations. *Sociology of Education*, *50*, 259–273.
- Blanton, H., Buunk, B. P., Gibbons, F. X., & Kuyper, H. (1999). When better-than-others compare upward: Choice of comparison and comparative evaluation as independent predictors of academic performance. *Journal of Personality and Social Psychology*, *76*, 420–430.
- Boyd, L., & Iverson, G. (1979). *Contextual analysis: Concepts and statistical techniques*. Belmont, CA: Wadsworth.
- Buckingham, J. T., & Alicke, M. D. (2002). The influence of individual versus aggregate social comparisons and the presence of others on self-evaluations. *Journal of Personality and Social Psychology*, *83*, 1117–1130.
- Butzer, B., & Kuiper, N. A. (2006). Relationships between the frequency of social comparisons and self-concept clarity, intolerance of uncertainty, anxiety, and depression. *Personality and Individual Differences*, *41*, 167–176.
- Chanal, J. P., Marsh, H. W., Sarrazin, P. G., & Bois, J. E. (2005). The big-fish–little-pond effect on gymnastics self-concept: Generalizability of social comparison effects to a physical setting. *Journal of Sport and Exercise Psychology*, *27*, 53–70.
- Cheng, R. W., & Lam, S. (2007). Self-construal and social comparison effects. *British Journal of Educational Psychology*, *77*(1), 197–211.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.
- Covington, M. V. (2001). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology*, *51*, 171–200.
- Dai, D. Y. (2004). How universal is the big-fish–little-pond effect? *American Psychologist*, *59*, 267–268.
- Dai, D. Y., & Rinn, A. N. (2008). The big-fish–little-pond effect: What do we know and where do we go from here? *Educational Psychological Review* (in press).
- Davis, J. A. (1966). The campus as a frog pond: An application of theory of relative deprivation to career decisions for college men. *American Journal of Sociology*, *72*, 17–31.
- Diener, E., & Fujita, F. (1997). Social comparison and subjective well-being. In B. P. Buunk & F. X. Gibbons (Eds.), *Health, coping, and well-being: Perspectives from social comparison theory* (pp. 329–358). Mahwah, NJ: Erlbaum.
- Eccles, J. S. (1983). Expectancies, values, and academic choice: Origins and changes. In J. Spence (Ed.), *Achievement and achievement motivation* (pp. 87–134). San Francisco: W. H. Freeman.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*, 117–140.
- Firebaugh, G. (1978). A rule for inferring individual-level relationships from aggregate data. *American Sociological Review*, *43*, 557–572.
- Goldstein, H. (2003). *Multilevel statistical models*. London: Edward Arnold.
- Hattie, J. C. (2002). Classroom composition and peer effects. *International Journal of Educational Research*, *37*, 449–481.
- Helson, H. (1964). *Adaptation-level theory*. New York: Harper & Row.
- Huguet, P., Dumas, F., Monteil, J.-M., & Genestoux, N. (2001). Social comparison choices in the classroom: Further evidence for students' upward comparison tendency and its beneficial impact on performance. *European Journal of Social Psychology*, *31*, 557–578.

- Hyman, H. (1942). The psychology of subjective status. *Psychological Bulletin*, *39*, 473–474.
- James, W. (1890/1963). *The principles of psychology* (vol. 2). New York: Holt.
- Kulik, C. L. (1985). *Effects of inter-class ability grouping on achievement and self-esteem*. Paper presented at the 1985 Annual Meeting of the American Psychological Association, Los Angeles.
- Kulik, C. L., & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal*, *21*, 799–806.
- Kulik, J. A., & Kulik, C.-L. C. (1991). Ability grouping and gifted students. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (pp. 178–196). Boston, MA: Allyn and Bacon.
- Kulik, J. A., & Kulik, C.-L. C. (1997). Ability grouping. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (pp. 230–242, 2nd ed.). Boston, MA: Allyn and Bacon.
- Lüdtke, O., Köller, O., Marsh, H., & Trautwein, U. (2005). Teacher frame of reference and the big-fish–little-pond effect. *Contemporary Educational Psychology*, *30*, 263–285.
- Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2007). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. <<http://www.cmm.bristol.ac.uk/research/Lemma/index.shtml>>.
- 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* (in press).
- Major, B., Testa, M., & Bylsma, W. H. (1991). Responses to upward and downward social comparisons: The impact of esteem-relevance and perceived control. In J. Suls & T. A. Wills (Eds.), *Social comparison: Contemporary theory and research* (pp. 237–260). New Jersey: Lawrence Erlbaum.
- Marsh, H. W. (1974). *Judgmental anchoring: Stimulus and response variables*. Unpublished doctoral dissertation, University of California, Los Angeles.
- Marsh, H. W. (1983). Neutral-point anchoring of personality-trait words. *American Journal of Psychology*, *96*, 513–526.
- Marsh, H. W. (1984a). Self-concept: The application of a frame of reference model to explain paradoxical results. *Australian Journal of Education*, *28*, 165–181.
- Marsh, H. W. (1984b). Self-concept, social comparison, and ability grouping: A reply to Kulik and Kulik. *American Educational Research Journal*, *21*(4), 799–806.
- Marsh, H. W. (1987). The big-fish–little-pond effect on academic self-concept. *Journal of Educational Psychology*, *79*, 280–295.
- Marsh, H. W. (1990). A multidimensional, hierarchical self-concept: Theoretical and empirical justification. *Educational Psychology Review*, *2*, 77–172.
- Marsh, H. W. (1991). The failure of high ability high schools to deliver academic benefits: The importance of academic self-concept and educational aspirations. *American Educational Research Journal*, *28*, 445–480.
- Marsh, H. W. (1993). Academic self-concept: Theory, measurement and research. In J. Suls (Ed.), *Psychological perspectives on the self* (vol. 4, pp. 59–98). Hillsdale, NJ: Erlbaum.
- Marsh, H. W. (1994). Using the national educational longitudinal study of 1988 to evaluate theoretical models of self-concept: The self-description questionnaire. *Journal of Educational Psychology*, *86*, 439–456.
- Marsh, H. W. (2005a). Big fish little pond effect on academic self-concept. *German Journal of Educational Psychology*, *19*, 119–128.
- Marsh, H. W. (2005b). Big fish little pond effect on academic self-concept: A reply to responses. *German Journal of Educational Psychology*, *19*, 141–144.
- Marsh, H. W. (2007). *Self-concept theory, measurement and research into practice: The role of self concept in educational psychology—25th Vernon–Wall lecture series*. London: British Psychological Society.
- Marsh, H. W., Chessor, D., Craven, R. G., & Roche, L. (1995). The effects of gifted and talented programs on academic self-concept: The big fish strikes again. *American Educational Research Journal*, *32*, 285–319.
- Marsh, H. W., & Craven, R. (1997). Academic self-concept: Beyond the dustbowl. In G. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement, and adjustment* (pp. 131–198). Orlando, FL: Academic.
- Marsh, H. W., & Craven, R. (2002). The pivotal role of frames of reference in ASC formation: The big fish little pond effect. In F. Pajares, & T. Urdan (Eds.) *Adolescence and education* (vol. II, pp. 83–123). Greenwich, CT: Information Age.
- Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, *1*, 133–163.
- Marsh, H. W., & Hau, K. T. (2003). Big fish little pond effect on academic self-concept: A crosscultural (26 country) test of the negative effects of academically selective schools. *American Psychologist*, *58*, 364–376.

- Marsh, H. W., & Hau, K.-T. (2007). Applications of latent-variable models in educational psychology: The need for methodological–substantive synergies. *Contemporary Educational Psychology*, *32*, 151–171.
- Marsh, H. W., Hau, K., & Craven, R. G. (2004). The big-fish–little-pond effect stands up to scrutiny. *American Psychologist*, *59*, 268–271.
- Marsh, H. W., Köller, O., & Baumert, J. (2001). Reunification of East and West German school systems: Longitudinal multilevel modeling study of the big fish little pond effect on academic self-concept. *American Educational Research Journal*, *38*, 321–350.
- Marsh, H. W., Kong, C.-K., & Hau, K.-T. (2000). Longitudinal multilevel models of the big-fish–little-pond effect on academic self-concept: Counterbalancing contrast and reflected-glory effects in Hong Kong schools. *Journal of Personality & Social Psychology*, *78*, 337–349.
- Marsh, H. W., & Parducci, A. (1978). Natural anchoring at the neutral point of category rating scales. *Journal of Experimental Social Psychology*, *14*, 193–204.
- Marsh, H. W., & O'Mara, A. J. (2008). *Have researchers underestimated the big-fish-little-pond-effect? Development of long-term total negative effects of school-average ability on diverse educational outcomes*. SELF Research Centre, University of Oxford.
- Marsh, H. W., & Parker, J. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, *47*, 213–231.
- Marsh, H. W., & Peart, N. (1988). Competitive and cooperative physical fitness training programs for girls: Effects on physical fitness and on multidimensional self-concepts. *Journal of Sport and Exercise Psychology*, *10*, 390–407.
- Marsh, H. W., & Rowe, K. J. (1996). The negative effects of school-average ability on academic self-concept—An application of multilevel modeling. *Australian Journal of Education*, *40*, 65–87.
- Marsh, H. W., Smith, I. D., Barnes, J., & Butler, S. (1983). Self-concept: Reliability, stability, dimensionality, validity and the measurement of change. *Journal of Educational Psychology*, *75*, 772–790.
- Marsh, H. W., Trautwein, U., Lüdtke, O., & Baumert, J. (2007). The big-fish–little-pond effect: Persistent negative effects of selective high schools on self-concept after graduation. *American Educational Research Journal*, *44*(3), 631–669.
- Marsh, H. W., Trautwein, U., Lüdtke, O., & Köller, O. (2008). Social comparison and big-fish–little-pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology* (in press).
- 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*, 297–416.
- Meyer, J. W. (1970). High school effects on college intentions. *American Journal of Sociology*, *76*, 59–70.
- Morse, S., & Gergen, K. J. (1970). Social comparison, self-consistency, and the concept of self. *Journal of Personality & Social Psychology*, *16*, 148–156.
- Parducci, A. (1995). *Happiness, pleasure, and judgment: The contextual theory and its applications*. Mahwah, NJ: Erlbaum.
- Parducci, A., Perrett, D., & Marsh, H. W. (1969). Assimilation and contrast as range frequency effects of anchors. *Journal of Experimental Psychology*, *81*, 281–288.
- 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.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks: Sage.
- Rogers, S. (1941). The anchoring of absolute judgments. *Archives of Psychology*, *37*(261), 1–42.
- Schwarzer, R., Lange, B., & Jerusalem, M. (1982). Selbstkonzeptentwicklung nach einem Bezugsgruppenwechsel [Self-concept development after a reference-group change]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *14*, 125–140.
- Seaton, M. (2007). The big-fish–little-pond effect under the grill: Tests of its universality, a search for moderators, and the role of social comparison. Unpublished doctoral dissertation, University of Western Sydney, Australia.
- Seaton, M., Craven, R. G., & Marsh, H. W. (2008a). East meets west: An examination of the bigfish-little-pond effect in western and non-western countries. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *International advances in self research* (vol. 3, pp. 353–371). Greenwich, CT: Information Age Press.
- Seaton, M., Marsh, H., & Craven, R. G. (2008b). *Earning its place as a pan-human theory: Universality of the big-fish–little-pond effect (BFLPE) across 41 culturally and economically diverse countries*. SELF Research Centre, University of Western Sydney, Sydney, Australia.

- Seaton, M., Marsh, H. W., Dumas, F., Huguet, P., Monteil, J.-M., Régner, I., *et al.* (2008c). In search of the big-fish: Investigating the coexistence of the big-fish–little-pond effect with the positive effects of upward comparison. *British Journal of Social Psychology*, *47*, 73–33.
- Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology*, No. 187.
- Sherif, M., & Sherif, C. W. (1969). *Social psychology*. New York: Harper & Row.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.
- Stapel, D. A., & Koomen, W. (2001). I, we, and the effects of others on me: How self-construal level moderates social comparison effects. *Journal of Personality & Social Psychology*, *80*, 766–781.
- Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., & Williams, R. M. (1949). *The American soldier: Adjustments during army life* (vol. 1). Princeton: Princeton University Press.
- Suls, J. M. (1977). Social comparison theory and research: An overview from 1954. In J. M. Suls & R. L. Miller (Eds.), *Social comparison processes: Theoretical and empirical perspectives* (pp. 1–20). Washington, DC: Hemisphere.
- Suls, J., & Wheeler, L. (2000). A selective history of classic and neo-social comparison theory. In J. Suls & L. Wheeler (Eds.), *Handbook of social comparison: Theory and research* (pp. 3–19). New York: Kluwer Academic/Plenum.
- Torgerson, C. J. (2003). *Systematic reviews*. London: Continuum Research Methods.
- Trautwein, U., Gerlach, E., & Lüdtke, O. (2008a). Athletic classmates, physical self-concept, and free-time physical activity: A longitudinal study of frame of reference effects. *Journal of Educational Psychology* (in press).
- Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology*, *98*, 788–806.
- Trautwein, U., Lüdtke, O., Marsh, H. W., & Nagy, G. (2008b). *Within-school social comparison: Perceived class prestige predicts academic self-concept*. Berlin: Max Planck Institute for Human Development.
- Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen & K. Elliot (Eds.), *But what does it mean? The use of effect sizes in educational research* (pp. 55–66). London: National Foundation for Educational Research.
- Upshaw, H. S. (1969). The personal reference scale: An approach to social judgment. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (vol. 4, pp. 315–370). New York: Academic.
- Volkman, J. (1951). Scales of judgment and their implications for social psychology. In J. H. Loher & M. Sherif (Eds.), *Social psychology at the crossroads* (pp. 273–294). New York: Harper.
- Wedell, D. H., & Parducci, A. (2000). Social comparison: Lessons from basic research on judgment. In J. Suls & L. Wheeler (Eds.), *Handbook of social comparison: Theory and research. The Plenum series in social/clinical psychology* (pp. 223–252). Dordrecht, Netherlands: Kluwer Academic.
- Wheeler, L., & Suls, J. (2004). Social comparison and self-evaluations of competence. In A. Elliot & C. Dweck (Eds.), *Handbook of competence and motivation*. New York: Guilford.
- Zeidner, M., & Schleyer, J. (1998). The big-fish–little-pond effect for academic self-concept, test anxiety, and school grades in gifted children. *Contemporary Educational Psychology*, *24*, 305–329.