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

Marsh, H., Morin, A., & Parker, P. (2015). Physical self-concept changes in a selective sport high school: A longitudinal cohort-sequence analysis of the big-fish-little-pond effect. *37*(2), 150-163.

Physical Self-Concept Changes in a Selective Sport High School: A Longitudinal Cohort-Sequence Analysis of the Big-Fish-Little-Pond Effect

Elite athletes and nonathletes (N = 1,268) attending the same selective sport high school (4 high school age cohorts, grades 7–10, mean ages varying from 10.9 to 14.1) completed the same physical self-concept instrument 4 times over a 2-year period (multiple waves). We introduce a latent cohort-sequence analysis that provides a stronger basis for assessing developmental stability/change than either cross-sectional (multicohort, single occasion) or longitudinal (single-cohort, multiple occasion) designs, allowing us to evaluate latent means across 10 waves spanning a 5-year period (grades 7–11), although each participant contributed data for only 4 waves, spanning 2 of the 5 years. Consistent with the frame-of-reference effects embodied in the big-fish-little-pond effect (BFLPE), physical self-concepts at the start of high school were much higher for elite athletes than for nonathlete classmates, but the differences declined over time so that by the end of high school there were no differences in the 2 groups. Gender differences in favor of males had a negative linear and quadratic trajectory over time, but the consistently smaller gender differences for athletes than for nonathletes did not vary with time.

Keywords: physical self-concept, big-fish-little-pond effect (BFLPE), cohort-sequence design, growth models

The present investigation is a methodological substantive synergy (Marsh & Hau, 2007; Seaton & Marsh, 2013) in which we introduce new and evolving methodology and statistical models likely to be useful to sport/exercise psychologists in evaluating substantively important issues. Methodologically, we introduce a latent variable model of multiwave-multicohort data based on cohort-sequence data. This allows us simultaneously to evaluate both measurement invariance and the extent to which measurement properties generalize over multiple groups (elite-athlete vs. nonathlete; male vs. female; multiple age cohort groups) and over time (i.e., observing the same group of participants on multiple occasions). Our substantive focus is on stability/ change in physical self-concept for male and female elite athletes and nonathletes over the potentially turbulent adolescent period. More specifically we consider elite athletes and nonathletes attending the same selective sport high school (four age cohorts, grades 7–10 in high school in the first year of the study) who completed the same physical self-concept measure four times over a 2-year period (multiple waves). We also evaluate predictions from the big-fish-little-pond effect (BFLPE), which posits that the initially large differences in physical self-concept in favor of elite athletes measured at the onset of attending a selective sport high school will decline sharply over time.

Substantive Focus on the Development of Physical Self-Concept

There is a revolution sweeping psychology, one that emphasizes a positive psychology and focuses on how healthy, normal and exceptional individuals can get the most from life (e.g., Seligman & Csikszentmihalyi, 2000; Vallerand et al., 2003). Consistently with this emphasis, a positive self-concept is valued as a desirable outcome in many disciplines of psychology (Marsh, 2007b; Harter, 2012; Hattie, 1992). Marsh (2007b) argues that self-concept is also an important mediating factor that facilitates the attainment of other desirable psychological and behavioral outcomes; that the need to think and feel positively about oneself, and the benefits of positive cognitions on choice, planning, and subsequent accomplishments, transcend traditional disciplinary barriers, and are central to goals in many social policy areas.

In their classic review of self-concept research, Shavelson, Hubner and Stanton (1976) examined existing definitions of self-concept, developed a construct definition of self-concept on the basis of previous theoretical work, and posited the structural model of self-concept, which has been a cornerstone of subsequent self-concept research. Based on an amalgamation of contemporary theories, they defined self-concept as the way a person perceives him- or herself, based on their experience and interpretation of their environment. Self-concept is particularly influenced by the views of significant others, by reinforcement, and attributions for one’s own behavior. Self-concept affects the way one behaves, and this behavior in turn subsequently affects one’s self-concept. Shavelson et al. (1976) emphasized that self-concept is important not only as an outcome, but also as a mediating factor on other important outcomes. Of particular relevance, they noted that self-concept can be based on an absolute ideal, on a personal standard, on a relative standard produced out of comparisons with others, or on the expectations of significant others.

Consistent with this focus on self-concept in many disciplines, in sport/exercise psychology, physical selfconcept is frequently posited both as a valued outcome variable and as a mediating variable that facilitates the attainment of other desirable outcomes, such as physical skills, health-related physical fitness, physical activity, exercise adherence in nonelite settings (Dishman, O’Connor & Tomporowski, 2013; Fox & Corbin, 1989; Marsh, 1997; Marsh & Cheng, 2012; Marsh, Gerlach, Trautwein, Lüdtke & Brettschneider, 2007; Standage, Gillison, Ntoumanis & Treasure, 2012; Spray, Warburton & Stebbings, 2013), and improved performance in elite sports (Marsh, Perry, Horsely & Roche, 1995).

Here we focus on the stability/change over time of physical self-concept during the potentially turbulent adolescent period (Dusek & Flaherty, 1981), a period of important biopsychosocial changes that are critical for the positive development of boys and girls, and of athletes and nonathletes. In the present investigation, elite athletes and nonathletes attending the same selective sport high school (four age cohorts, grades 7–10 in high school) completed the same physical self-concept instrument four times over a 2-year period (multiple waves). Analyzed through new and evolving statistical models, these results provide a unique opportunity to assess stability/change in male and female athletes and nonathletes over the adolescent period. We also evaluate predictions from the big-fish-little-pond effect (BFLPE), widely supported in academic self-concept research, which posits that the initially large differences in physical self-concept in favor of athletes measured at the start of attending a selective sport high school will decline sharply over time.

Self-Concept: Age, Gender and Athlete Group Differences

Predictions about how self-concept develops with age have been developed from a variety of theoretical perspectives. Marsh (1989; 2007b; Marsh, Craven & Debus, 1991) proposed that the self-concepts of very young children are consistently high but that with increasing life experience, children learn their relative strengths and weaknesses, so that mean levels of self-concept decline, multiple dimensions of self-concept become more differentiated, and self-concepts become more highly correlated with external indicators of competence (e.g., skills, accomplishments, and self-concepts inferred by significant others). Eccles, Wigfield, Harold, and Blumfeld (1993; Fredricks, & Eccles, 2002; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Wigfield & Cambria, 2010) similarly proposed that the declines in mean levels of self-concept, particularly during the preadolescent and early-adolescent periods, reflected an optimistic bias in young children and increased accuracy in responses as they grow older. Based on a large empirical study, Marsh (1989) reported that there was a reasonably consistent pattern of selfconcepts declining from a young age into at least early adolescence, leveling out, and then increasing at least into early adulthood. In related research, Harter (2012) proposed that self-concept becomes increasingly abstract with age, shifting from concrete descriptions of behavior in early childhood, to traitlike psychological constructs (e.g., popular, smart, good looking) in middle childhood, to more abstract constructs during adolescence.

Historically, gender differences in self-concept typically have focused on small differences in self-esteem or total self-concept favoring boys (e.g., Feingold, 1994). Marsh (1989, 2007b), however, demonstrated that these small differences in global components of self-concept reflect larger, counterbalancing gender differences in specific components of self-concept. The gender differences in more specific facets of self-concept tend to be consistent with traditional gender stereotypes: (a) boys had higher Physical Ability, Appearance, Math, Emotional Stability, Problem Solving, and Esteem self-concepts; (b) girls had higher Verbal/Reading, School, Honesty/Trustworthiness, and Religion/Spiritual Values self-concepts; and (c) no gender differences for Parents self-concepts. For preadolescent, early-adolescent, and late-adolescent age ranges and into early adulthood, the largest gender differences were the higher scores for boys in the physical components of self-concept (also see Beasley, 2013).

In the sport/activity domain, Marsh, Hey, Roche and Perry (1997; also see Marsh, Perry et al., 1995) compared the self-concept responses of a broad cross-section of athletes from the Australian Institute of Sport with those in the normative archive group for the Self-Description Questionnaire III. Not surprisingly, the largest difference was in the substantially higher Physical Ability self-concepts of the athletes (also see Luszczynska & Abraham, 2012). There were smaller differences favoring athletes on the social components of self-concepts (Same Sex, Opposite Sex, and Parent Relationships) and on Global Self-Esteem. Although gender differences in the general population tended to be mirrored in the athlete group, there were notable exceptions. Particularly relevant to the present investigation, gender difference in the Physical Ability self-concept scale was clearly the largest difference in the general population, although it was not statistically significant for the athlete group.

Big-Fish-Little-Pond Effects: Frame-of-Reference Effects

Psychologists from the time of William James (1890/1963) have recognized that objective accomplishments are evaluated in relation to frames of reference. James noted 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). The same objective accomplishment can lead to quite different self-concepts, depending on the frames of reference or standards of comparison against which individuals evaluate themselves. Marsh (1984; see also Marsh & Parker, 1984; Marsh et al., 2008) proposed the BFLPE to encapsulate frame-of-reference effects that are based on an integration of theoretical models and empirical research from diverse disciplines (e.g., relative deprivation theory, social comparison theory, psychophysical judgment, social judgment; see Marsh, 1984; Marsh et al., 2008). In the BFLPE model, students are hypothesized to compare their abilities with the abilities of their classmates, and to use this social comparison impression as one basis for forming their own self-concept; individual ability is positively related to academic self-concept (the more able I am, the higher my self-concept) but class- and school-average abilities have a negative effect on self-concept (the more able my classmates, the lower my academic self-concept). Hence, self-concept depends not only on a student's own accomplishments but also on those of their classmates.

BFLPE Generalizability. Across diverse samples, ages, instruments, and designs, there is extensive empirical support for the BFLPE in the academic domain (Marsh et al., 2008; Nagengast & Marsh, 2012), making it one of the most robust effects in psychological research. Particularly important, in support of its generality, is the set of studies based on successive data collections of the OECD-PISA (Organization for Economic Cooperation and Development Program for International Student Assessment). Summarizing the three BFLPE-PISA studies, Nagengast and Marsh (2012) reported that the effect of school-average achievement was negative in all but one of the 123 country-samples across the three studies, and significantly so in 114 samples. The average effect size across all 123 samples was –.223.

Local Dominance Effects. A critical issue in BFLPE research is conceptualization of the appropriate frame of reference. BFLPE studies typically have defined the frame of reference as either the school as a whole or the local classroom. However, based on social psychology laboratory studies, Zell and Alicke (2009; see also Alicke, Zell, & Bloom, 2010) experimentally manipulated the frame of reference in relation to feedback given to participants about how their performances compared with others. They pitted idiosyncratic local against more general broadly representative comparison standards. Participants consistently used the most local comparison information available to them, even when they were told that the local comparison was not representative of the broader population and they were provided with more appropriate normative comparison standards. Integrating the local dominance effect and the BFLPE, Marsh, Kuyper, Morin, Parker, and Seaton (2014) juxtaposed frame-of-reference effects associated with school and class in the Dutch school system. Although they found BFLPEs in relation to both class and school, when considered together, the classlevel effects remained substantial while the school level effects were no longer significant. Also consistent with local dominance, self-concepts were largely determined by comparisons with students in their own class, but not by comparisons with students in other classes or schools.

Juxtaposition of the Contrast and Assimilation Effects. Based on the self-concept theory underpinning the BFLPE (Marsh, 1984; 1987; Marsh & Craven, 2002; Marsh et al., 2008), there is a juxtaposition of assimilation (positive social comparison, sometimes referred to as reflected glory; i.e., if I am good enough to be selected to be in this program, I must be pretty good), leading to more positive self-concepts, and contrast (negative social comparison, here referred to as a negative BFLPE; i.e., relative to all these other elite athletes, maybe I am not as good as I thought), leading to more negative self-concepts. Although assimilation effects have a clear theoretical basis, these effects have been largely implicit and elusive in BFLPE studies (Marsh, Kong & Hau, 2000; Marsh et al., 2014; Trautwein, Lüdtke, Marsh & Nagy, 2009). Observed BFLPEs are likely to represent the counterbalancing, net effects of these two opposing processes. The implication is that an assimilation (positive social comparison) effect may be operating, even though its effects are not as large as the contrast (negative BFLPE) effects, but research has been only partially successful in disentangling the two effects in BFLPE studies.

Marsh (1984; Marsh, Köller, Baumert, 2001) further posited that there may be temporal differences in the development of the counter-balancing assimilation (positive social comparison, or reflected glory) effects and contrast (negative social comparison, or negative BFLPE) effects. Thus, for example, students who have just learned of their selection into a special program may bask in the glory of this accolade without experiencing the need to evaluate themselves in relation to their new classmates, whom they have not yet met—an assimilation effect. After starting the new program, however, the relative accomplishments of different students will become more salient with increasing levels of explicit or implicit performance feedback, and social comparison, encouraging students to reestablish their pecking order within this new context—a contrast effect, or negative BFLPE. Although there is limited support for this juxtaposition of assimilation and contrast in relation to temporal ordering, the results are consistent with a number of studies showing that the BFLPE grows more negative, the more years a student spends in the same school (Marsh, 1991; Marsh et al., 2000, 2001; Marsh & Craven 2002; Marsh, Trautwein, Lüdtke, Baumert, & Köller, 2007) and has lasting effects on a variety of important outcomes even after students leave high school (Marsh, 1991; Marsh & O'Mara, 2010; Marsh, Trautwein, et al. 2007). Of particular relevance, Marsh et al. (2001) found that the BFLPEs became substantially more negative during the first year after reunification for East German students who had not previously experienced selective schools, compared with West German students who had previously attended selective schools for the 2 years before reunification. For East German students, the BFLPE was not evident at the start of the school year, grew larger but was still less than in West German students by the middle of the school year, and became 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 become evident.

The BFLPE in Sport/Exercise Settings. The comparative development of physical self-concept for athletes and nonathletes during this adolescent period is of general interest in that is both an important outcome in its own right and an important predictor of other important outcomes (Beasley, 2013; Fox & Corbin, 1989; Marsh, 1994; Marsh & Cheng, 2012; Marsh, Gerlach, et al., 2007; Marsh, Martin & Jackson, 2010; Trautwein, Gerlach, & Lüdtke, 2008). In the present investigation, we are specifically interested in the stability/change in physical self-concepts of athletes and nonathletes attending the same sports high school, and relating this to the BFLPE. For example, in their comparison of the physical selfconcepts of elite athletes, Marsh et al. (1995) noted that the same objective accomplishments can lead to very different self-concepts, depending on the different frames of reference. This led them to speculate that there might be a BFLPE affecting the physical self-concepts of elite athletes participating in selective sport programs. However, they recognized that "tests of this suggestion and an evaluation of implications if support is found are clearly beyond the scope of the present investigation, but provide an important area for further research" (Marsh et al., 1995, p. 80). Surprisingly, however, their call for more research on the BFLPE with elite athletes has apparently gone largely unheeded, as there have been no BFLPE studies in relation to elite athletes and few BFLPE studies in more general sport/exercise settings. In one notable exception, Marsh, Chanal, & Sarrazin (2006) evaluated the BFLPE based on responses by 405 students in 20 gymnastics classes. Consistent with a priori BFLPE predictions, gymnastics self-concept was positively predicted by individual gymnastics skills, but was negatively predicted by class-average gymnastics skills. The size of this BFLPE (i.e., the negative impact of the class-average skills) grew more negative during the 10-week training program (as participants had more exposure to the relative performances of others in their class), but did not vary as a function of gender, age, or initial gymnastics skills.

Noting the almost complete absence of the BFLPE in relation to physical self-concept, Beasley (2013) demonstrated that class-average physical ability in physical education classes had a negative effect on physical selfconcept. Although she found that girls had lower physical self-concepts than boys, contrary to predictions, class type (single-sex vs.

coeducational) did not moderate the BFLPE. Trautwein et al. (2008) likewise demonstrated that class-average physical ability had negative effects on physical self-concept for a large sample of primary school children aged 9 and 10 years of age. Furthermore, the negative effect of class-average physical ability led to subsequent declines in out-of-school physical activity and was substantially mediated by the declines in physical self-concept. Whereas the Marsh et al. (2006), Beasley (2013), and Trautwein et al. (2008) studies are apparently the only direct tests of the BFLPE on physical self-concept in a sport/exercise setting with nonelite athletes, there is apparently no research based on elite athletes, for whom the BFLPE might be expected to be particularly important: this is the focus of the present investigation.

Methodological Focus on Cohort Sequential Designs: Latent Longitudinal Models of Stability/ Change Across Multiple Cohorts and Multiple Occasions

Tests of measurement invariance evaluate the extent to which measurement properties generalize over multiple groups, situations, or occasions (Marsh, 2007a). Of particular importance for sport/exercise research is the evaluation of differences across multiple groups (e.g., athlete vs. nonathlete; male vs. female; age groups; treatment vs. control) or over time (i.e., observing the same group of participants on multiple occasions, perhaps before and after an intervention). The need for rigorous tests of whether the underlying factor structure is the same for different groups or occasions is often ignored in sport/ exercise research. Indeed, unless the underlying factors are measuring the same construct in the same way, and the measurements themselves are operating in the same manner across groups or time, the comparison of parameter estimates is potentially invalid. For example, if group or longitudinal differences in latent means vary substantially for different items used to infer a construct, in a manner unrelated to respondents' true levels on the latent construct, then the observed differences might be idiosyncratic to the particular items used. From this perspective, it is important to be able to evaluate the full measurement invariance of participants' responses.

As argued by Marsh (1998; Marsh, Craven & Debus, 1998; Parker, Marsh, Morin, Seaton & Van Zanden, 2014; also see Nesselroade & Baltes, 1979), a multiwave-multicohort design often provides a stronger basis for evaluating developmental differences than either cross-sectional comparisons based on many age cohorts or longitudinal comparisons of multiple waves for a single age cohort. Even though developmental researchers often extol the virtues of true longitudinal designs over cross-sectional designs, ultimately, support for the generality of developmental effects requires convergence of results across multiple approaches. Hence, the juxtaposition of longitudinal and cross-sectional approaches to developmental change within the same study provides an important basis for cross-validating results based on each approach in a way that would not be possible in longitudinal designs based on a single cohort or in crosssectional designs based on multiple cohorts.

In our study we collected four waves of data, 6 months apart, for each of four age cohorts. This cohortsequential design provided a total of 10 waves of data covering an entire 5-year period, in which the multiple waves of data overlapped for each successive cohort (see Figure 1 for structure), even though each individual participant provided data for only 2 of the 5 years. This cohort-sequence design provides sport/exercise researchers a feasible, cost-effective means to explore growth over the course of an entire developmental period (see Brodbeck, Bachmann, Croudace, & Brown, 2013; Enders, 2010; Graham, 2012). In the present investigation we extend the traditional manifest approach to be fully latent, thus allowing us to test the assumption of psychometric invariance over time and cohorts that underpins tests of means differences—latent or manifest.

The Present Investigation

In the present investigation we measured physical selfconcept four times over a 2-year period (multiple waves) for students (482 elite athletes, 786 nonathletes) in Years 7–10 in high school (multiple age cohorts). Implementing our latent cohort-sequence model provided a unique opportunity to evaluate how the physical self-concepts of elite athletes develop/change over the 5-year period, starting from the time when they are first selected to participate in a highly selective sporting program. With respect to the self-concept theory underpinning the BFLPE (Marsh, 1984; Marsh et al., 2008), there is a juxtaposition of assimilation, leading to more positive self-concepts, and contrast, leading to more negative self-concepts. Based on theory and research, we posit a temporal feature to this juxtaposition in which positive assimilation effects dominate when participants are first selected to be in an elite athlete program, but contrast effects become increasingly strong over time as their frame of reference shifts to other elite athletes in the same program (Marsh, 2007b; Marsh et al., 2001, 2008). On this theoretical basis we offer the following research hypotheses:

1. BFLPE Predictions: elite-athlete/nonathlete differences:

a. Elite athletes will have substantially higher physical self-concepts than nonathletes when they first start in the elite athlete program in Year 7.

b. However, consistent with predictions based on the BFLPE, the size of the difference in physical self-concept for the two (elite-athlete/nonathlete) groups will decline steadily over the 5-year period from the start of Year 7 to the end of Year 11.

2. Gender differences and the BFLPE

a. Boys will have substantially higher physical self-concepts than girls.

b. These gender differences will be larger for nonathletes than athletes.

c. However, we leave as research questions whether time interacts with gender (whether gender differences favoring boys grow larger or smaller with age) and the possibility of group-by-gender interaction (whether the decline over time in elite-athlete/nonathlete differences in support of the BFLPE varies as a function of gender).

Year and Group		Year 7 Cohort 1	Year 8 Cohort 2	Year 9 Cohort 3	Year 10 Cohort 4
Year 7	M1	M1			
	M2	M2			
Year 8	M3	M3	M3		
	M4	M4	M4		
Year 9	M5		M5	M5	
	M6		M6	M6	
Year 10	M7			M7	M7
	M8			M8	M8
Year 11	M9				M9
	M10				M10

Figure 1 — Cohort sequential design with four cohorts (year in school groups first tested in Years 7–10) and four waves of data for each cohort. Gray squares = Collected data. M1–M10 are 10 latent means that span the 5-year period. White square = Missing by design. Means in each box (solid black rectangles) are matching means based on two different cohorts. Estimates of M3–M8 are each based on results from two cohorts, while those for M1–M2 and M9–M10 are based on a single cohort.

Methods

Sample and Materials

The data were gathered through a larger study funded by the Australian Research Council. Study procedures and assessments were approved by ethics boards at the University of Western Sydney and the NSW Department of Education. Informed parental consent and participant assent were obtained before participation in the study. Participants (N = 1,268) were students in grades 7 (N = 302, mean age = 10.9), 8 (N = 340, mean age = 12.0), 9 (N = 317, mean age = 13.1), and 10 (N = 309, mean age = 14.1) at one of the most prestigious sports high schools in Australia. Each year, elite athlete students compete to enrol in most major sports (e.g., basketball, softball, rugby league, soccer, baseball, swimming, track and field, dance aerobics, cricket, and netball). The school also admits students from the local catchment area who typically do not participate in the elite athletic program. There were 478 (38%) elite athlete students (57% males, 43% females, mean age = 13.1 years) and 790 (62%) nonathlete students (53% males, 47% females, mean age = 13.3 years).

All elite athletes and nonathletes attending the same selective sport high school (four age cohorts, grades 7–10 in high school) completed the same physical self-concept instrument (the global physical scale from the Physical Self-Description Questionnaire; Marsh, 1996a, b; Marsh et al., 2010; Marsh, Richards, Johnson, Roche, & Tremayne, 1994) administered on four occasions (twice a year for 2 years, approximately 6 months apart; see Figure 1). Materials were administered by classroom teachers to intact classes of no more than 30 students, under the supervision of a trained school counselor, according to written instructions. The global physical self-concept scale was selected as being the most appropriate PSDQ scale in terms of being broadly representative, suitable for this sample (see Marsh, 1996a, b; 1998; Marsh & Cheng, 2012; Marsh et al., 2010), appropriate in relation to testing the BFLPE, and consistent with how physical self-concept is typically measured in other research. The six items in this physical self-concept scale are: I am satisfied with the kind of person I am physically; Physically, I am happy with myself; I feel good about the way I look and what I can do physically; Physically I feel good about myself; I feel good about who I am and what I can do physically; I feel good about who I am physically.

Statistical Analyses

For all models, physical self-concepts were specified as latent variables estimated from multiple items. This required relatively complex identification constraints. In most models we used a nonarbitrary metric for factor loadings and item intercepts, allowing the results to be interpreted according to the original 6-point Likert scale (see Little, Slegers & Card, 2006). As is typical in large longitudinal field studies, a substantial portion (23%) of the sample had missing data for at least one of the four occasions, due primarily either to student absence or the provision of an illegible (or fictitious) name on at least one of the four occasions. All models were fitted using the robust maximum likelihood estimator (MLR) available in Mplus 7.2, in conjunction with multiple imputation procedures for handling missing data based on unrestricted model-independent covariance imputation, which is the default in Mplus. The multiple imputation was based on a data file in which each student had four waves of data, so that imputation was used to fill in missing values within the four waves actually completed by each cohort of students, but not using the data from waves not completed by the cohort (i.e., imputation for the Year 7 cohort imputed data for the four waves of data in Years 7 and 8, but not for Years 9 and 10). The analyses were done on five imputed data sets and results were combined automatically by Mplus using the Rubin (1987; Schafer, 1997) strategy to obtain unbiased parameter estimates, standard errors, and goodness-of-fit statistics.

Based on the cohort-sequence model of these longitudinal data, we estimated 10 latent means (i.e., 2 means per year for each of the 5 years covered by this design). To evaluate the nature of change over time, we then fitted a latent growth model based on traditional orthogonal polynomial contrasts (e.g., Cohen, West & Aiken, 2014; Von Eye & Schuster, 1998) to estimate polynomial (i.e., linear, quadratic, cubic, etc.) components. Orthogonal polynomials have two defining characteristics: they sum to 0 for each component and are mutually independent for pairs of components (Cohen et al., 2014). Because the polynomial components are uncorrelated, they are easy to evaluate and interpret. Although orthogonal polynomial contrasts are often defined in terms of integer values, interpretability is enhanced by using contrast coefficients that are normalized, such that the sum of squared coefficients for each contrast sums to 1.0 (e.g., Maxwell & Delaney, 2004). Although familiar to most researchers who use contrasts in analyses of manifest models (e.g., polynomial contrasts with ANOVA or multiple regression) using standard statistical packages such as SPSS (where polynomial contrasts are the default for repeated-measures ANOVAs), here the polynomial functions are fitted to latent means as part of the same analyses used to estimate the latent factor structure. In combination with grouping variables (gender: male vs. female; athlete group: elite-athlete vs. nonathlete) it is then possible to test interactions between each of the growth curve components, and various combinations of these grouping variables (e.g., does change over time differ for males and females).

Preliminary Analyses: Invariance Over Time and Cohorts

Longitudinal models that evaluate development/change of latent means across multiple waves require strong/ scalar invariance in which factor loadings and item intercepts are equal across time waves. In a cohort sequential design, these parameters are assumed to be invariant across both time waves and cohorts (thus providing a direct test of possible historical/cohort effects; see Supplemental Materials for further discussion). Furthermore, to estimate the growth trajectories from all time waves and cohorts, cohort-sequential designs also assume that overlapping latent means are invariant across cohorts (e.g., T3 and T4 of the Year 7 cohort, when they are in Year 8, are the same as T1 and T2 of the Year 8 cohort; see boxes in Figure 1). Thus the model constrains loadings and item intercepts to be invariant over time (strong/scalar invariance), and the latent means for overlapping time points to be invariant across time points—providing an additional test of cohort-specific historical effects. This highly demanding pattern of invariance is an inherent assumption of cohort sequential designs for models involving latent means, but is rarely tested in applied research.

Here our focus is on the analysis of cohort sequence data in relation to substantive concerns about stability and change in physical self-concept over time. However, these analyses are based on critical assumptions about the invariance of responses over time and cohort that are complex in cohort-sequence designs. The preliminary analyses that underpin our research are presented in detail in the Supplemental Materials, along with the Mplus scripts used in the analysis of the results. In brief, these results show that the factor structure of the physical self-concept items was invariant over all combinations of the four cohorts (year in school) and the four waves of data (the four occasions each student completed the physical self-concept measure). More specifically, there was good support for factor loading (metric) invariance and item intercept (scalar) invariance. Furthermore, there was good support for the invariance of latent means for overlapping means in the cohort-sequence design (Figure 1). Thus, for example, the two estimates of M3 based on different cohorts (Wave 3, Year 7 cohort and Wave 1, Year 8 cohort) were constrained to be equal.

Next, we extended this latent cohort-sequence model to address the research hypotheses and questions posed earlier. It is of substantive interest to compare developmental stability/change in physical self concept over time as a function of gender, athletic group, and their interaction. This was accomplished by integrating the traditional MIMIC model (i.e., treating gender, athletic group, and their interaction as three covariates) with the final cohort-sequence model (with factor loading and intercepts invariant over time, and invariance of overlapping means in the cohort-sequence model described above; see Supplemental Materials for more detail). In a series of MIMIC models (see Supplemental Materials) we showed that intercepts associated with each of the MIMIC variables were invariant over all cohorts and occasions, and that the effects of the MIMIC variables were invariant within each of the six pairs of matching means (i.e., the matching means in boxes in Figure 1). Substantive interpretations discussed in the results are based on this final MIMIC model.

Results

An apparently unique aspect of the present investigation is that 38% of our adolescent sample were elite athletes selected through a highly competitive process, whereas the remaining 62% were nonathletes of similar ages attending the same school and taking mostly the same classes. From this perspective it is of substantive interest to compare developmental stability/change in physical self-concept over time as a function of gender for these two groups of students. Integrating the traditional MIMIC model (treating gender, athletic group, and their interaction as three covariates) with the cohort-sequence analyses (see Supplemental Materials) provides a substantive-methodological synergy in which we apply new and evolving statistical models to evaluate substantively important issues in relation to self-concept theory. Based on this MIMIC model (described in more detail in the Supplemental Materials), we estimated a latent growth model based on traditional orthogonal polynomial contrasts (e.g., Cohen et al., 2014) in combination with the "model constraint" function available in Mplus to evaluate more fully the latent means based on the cohort-sequence model and the effects of the MIMIC variables on latent physical self-concept factors over time.

We began with the total effects for each wave, averaged across gender and group. These are presented in the form of a table (Table 1) that highlights the cohort-sequence design underpinning these data. In Table 2 the same values are presented as a single column of means (the "Total" column in Table 2), along with tests of the orthogonal polynomial contrasts. These results show that, averaged across all groups, there are statistically significant linear (negative) and quadratic effects over time; from the start of Year 7 (the first two data waves for the Year 7 cohort) to the end of Year 11 (the last two waves of the Year 10 Cohort). Thus, the functional relation between time and physical self-concept includes both a linear and a quadratic component, such that both need to be considered simultaneously. Inspection of the latent means (Table 2; also see Figure 2) illustrates that there is a decline in physical self-concept from M1 (start of Year 7, M1 = 4.598) to M7 (start of Year 10, M7 = 4.260), followed by a small increase in the last three waves (end of Year 11, M10 = 4.352). The cubic and quartic polynomials are not statistically significant. This general pattern of results is consistent with other self-concept research covering an even wider age range, which shows that many areas of self-concept decline from early childhood, level off in middle adolescence, and increase through early adulthood (e.g., Marsh, 1989; see earlier discussion). However, because our primary interest is in stability/change in relation to gender, athlete group, and their interaction, we now extend the analyses to focus on these issues.

Table 1 Latent Factor Means (*SEs* in Parentheses) of Physical Self-Concept for Model M4–6b (Table 1 in Supplemental Materials): Complete Invariance of Factor Loadings and Intercepts Over Time and Cohort, Invariance of Matching Latent Means Over Cohort (in Boxes)

	Year 7–8 (Cohort 1)	Year 8–9 (Cohort 2)	Year 9–10 (Cohort 3)	Year 10–11 (Cohort 4)
M1-Yr7	W1 4.598(.095)			
M2-Yr7	W2 4.528(.100)			
M3-Yr8	W3 4.463(.086)	W1 4.463(.086)		
M4-Yr8	W4 4.365(.089)	W2 4.365(.089)		
M5-Yr9		W3 4.390(.090)	W1 4.390(.090)	
M6-Yr9		W4 4.358(.088)	W2 4.358(.088)	
M7-Yr10			W3 4.260(.083)	W1 4.260(.083)
M8-Yr10			W4 4.347(.083)	W2 4.347(.083)
M9-Yr11				W3 4.293(.095)
M10-Yr11				W4 4.352(.090)

Note. Cohort is the school year group for the first data collection. Wave 1–Wave 4 (W1–W4) are the four waves within each cohort. M1–M10 are the estimated latent means (matching means within boxes are constrained to be equal over cohort; see Figure 1). Standard errors (*SEs*) in parentheses provide a test of significance for each estimate (if the ratio of the estimate/*SE* is greater than 2, the difference is significant at $p < .05$).

The next three columns in Table 2 present the main effects of gender and athlete group and their interaction (also see Figure 2). For each of these three effects the "grand mean" represents the main effects of each variable averaged across the 10 waves (M1–M10 in Table 2) in our longitudinal design. The first statistical test associated with each of these variables is a test of the statistical significance of the main effects (e.g., group differences averaged over time). Next we evaluated developmental stability/change in these main effects; whether gender and group effects are longitudinally consistent across the 10 cells of our cohort-sequence design. In the language of ANOVA we were testing the time-by-gender, time-by-group, and time-by-gender-bygroup interactions—keeping in mind that these contrasts are applied to latent means based on our cohort-sequence design. We now evaluate these effects in relation to the a priori hypotheses.

Research Hypothesis 1: Elite-Athlete/ Nonathlete Differences and the BFLPE

Consistent with Research Hypothesis 1a, not surprisingly, there is a highly significant main effect of athlete group; averaged over time, athletes have substantially higher physical self-concepts than nonathletes. However, consistent with Research Hypothesis 1b, the athlete group differences vary substantially over time (start of Year 7 to the end of Year 11). Inspection of the polynomial contrasts shows that only the linear component of these changes over time is statistically significant; quadratic, cubic, and quartic components are not statistically significant.

As predicted, when elite athletes first join the elite sporting program (M1, at the start of Year 7) they have substantially more positive physical self-concepts than the corresponding Year 7 nonathletes attending the same high school. However, also consistent with a priori predictions based on the BFLPE, there is a substantial linear decline over time in the size of physical self-concept differences for the athlete and nonathlete groups. Inspection of the latent means representing the 10 waves in our cohort-sequence design (Table 2) indicates that group differences in favor of elite athletes are substantial at the start of high school (–.634 for M1 in Year 7) but decline rapidly, so that apparent group differences in favor of athletes in the final two occasions (M9 and M10 in Year 11 in Table 2; also see Figure 2) are not even statistically significant.

Table 2 Development as a Function of Athlete/Nonathlete Group and Gender: Multigroup (Four Cohorts) Cohort-Sequence Analysis With Gender, Group and Their Interaction as MIMIC Variables with Orthogonal Contrasts of Time Evaluated With Orthogonal (Model Constraint) Contrasts

		Main Effects		Interaction
Wave	Total *M* (*SE*)	Gender (F/M) *M* (*SE*)	Athlete (A/NA) *M* (*SE*)	Gender × Athlete *M* (*SE*)
M1-Yr7	4.598(.095)*	–.146(.107)*	.634(.098)*	.129(.211)
M2-Yr7	4.528(.100)*	–.146(.110)*	.460(.110)*	.176(.230)
M3-Yr8	4.463(.086)*	–.237(.075)*	.356(.074)*	.208(.140)
M4-Yr8	4.365(.089)*	–.433(.080)*	.427(.077)*	.226(.159)
M5-Yr9	4.390(.090)*	–.515(.079)*	.526(.073)*	.298(.156)
M6-Yr9	4.358(.088)*	–.516(.078)*	.321(.076)*	.428(.163)*
M7-Yr10	4.260(.083)*	–.690(.092)*	.237(.084)*	.102(.175)
M8-Yr10	4.347(.083)*	–.624(.086)*	.312(.085)*	.156(.178)
M9-Yr11	4.293(.095)*	–.447(.109)*	.253(.146)	.178(.283)
M10-Yr11	4.352(.090)*	–.494(.111)*	.198(.110)	.148(.244)
Summary of Model Contrast				
Mean	4.396(.080)*	–.425(.045)*	.373(.043)*	.205(.093)*
ES	–0.362	0.317	0.174	—
Linear	–.264(.067)*	–.437(.130)*	–.350(.137)*	–.017(.283)
ES	–0.225	–0.372	–0.298	–0.015
Quadratic	.145(.050)*	.289(.103)*	.032(.097)	–.164(.215)
ES	0.123	0.246	0.027	–0.139
Cubic	.007(.043)	.120(.092)	–.060(.087)	.052(.179)
ES	0.006	0.102	–.051	0.044
Quartic	.012(.041)	–.108(.081)	.097(.088)	.095(.196)
ES	0.01	–0.092	0.083	0.081

Note. M1–M10 are the estimated latent means for the 10 waves, two means for each of five school years (Yr7–Yr11; see Figure 1). *SE*s = standard errors. For each of the covariates (MIMIC variables: gender [F/M] = females – males difference; athlete [A/NA] = elite-athlete/nonathlete; and their interaction), the corresponding tests evaluate the grand mean and the nature of growth. The grand means are the mean of M1–M10 and provide an overall test of each covariate. Polynomial components test the nature of growth. Results are based on Model M4–6b in Table 1 in Supplemental Materials (also see Supplemental Materials for Mplus syntax).
*$p < .05$.

Research Hypothesis 2: Gender Differences for Elite Athletes and Nonathletes and the BFLPE Consistent with prediction 2a, there was a highly significant main effect of gender: averaged across all 10 waves (start of Year 7 to the end of Year 11) boys had significantly more positive physical self-concepts than girls. However, these gender differences varied significantly over time; there is a significantly negative linear effect, and a significantly positive quadratic effect; cubic and quartic components are not significant. Inspection of the latent means (Table 2; also see Figure 2) demonstrates that gender differences in favor of boys are smallest at the start of high school (–.146 for M1 and M2 in Year 7), increase over time (–.690 for M7 at the start of Year 10), and then become smaller (–.447 and –.494 for M9 and M10 in Year 11).

Consistent with Hypothesis 2b, there is a significant gender-by-group interaction, such that gender differences in physical self-concept are significantly smaller for athletes than nonathletes. However, there are no significant differences over time for the genderby-group interaction. Thus, gender differences in favor of males are smaller for athletes, but this difference is reasonably stable across the 10 waves spanning the period from the start of Year 7 to the end of Year 11.

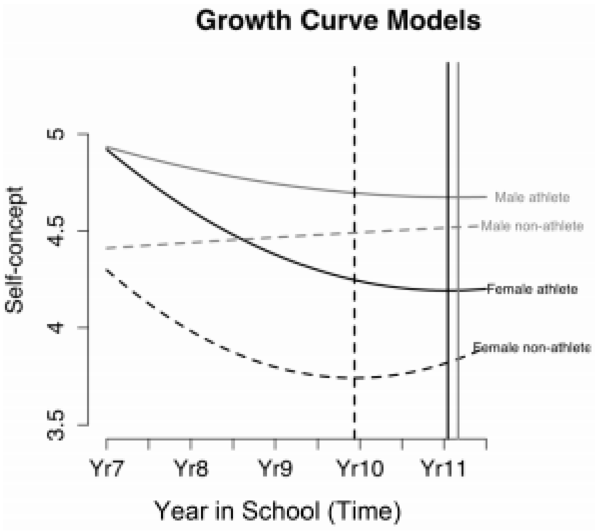

Figure 2 — Changes in Physical Self-concept over time: Plots of the polynomial growth curves for each of the four groups (male and female elite-athletes and nonathletes) over the 10 time waves over five school years (Yr7–Yr11) in the cohort-sequence design (see Figure 1 and Tables 1 and 2). Horizontal lines represent the "turning point" (vertex) for each of the quadratic functions (except for the male nonathlete group, where the function is linear, thus having no vertex).

Discussion

Substantive Findings and Implications

Our study provided an apparently unique opportunity to evaluate stability/change over the course of high school for elite athletes and nonathletes attending a selective sports high school. The temporal pattern of elite-athlete/ nonathlete differences is consistent with a priori hypotheses based on the well-established frame-of-reference effects posited by the BFLPE. In the BFLPE, students attending academically selective schools experience increasingly negative academic self-concepts over the course of high school, compared with equally able students who attend nonselective schools. Extrapolating from this research in the academic domain, we speculated that at the start of high school, elite athletes form their physical self-concepts in relation to a broadly defined cross-section of previous classmates, resulting in very positive physical self-concepts. However, over the high school years at a highly selective sports high school, these elite athletes increasingly tend to form their physical self-concepts in relation to other elite athletes attending the same high school. Hence, we speculated that the decline in physical self-concept reflects the increasingly more demanding frame of reference against which elite athletes evaluate themselves.

We were also interested in gender differences in physical self-concept. Consistent with a priori hypotheses, girls, not surprisingly, had substantially lower physical self-concepts than boys, but these gender differences were substantially smaller for athletes than nonathletes. However, the sizes of the gender-by-group interactions did not vary significantly over time. Thus, the decline in physical self-concepts for athletes relative to nonathletes—a BFLPE—was consistent for boys and girls. In this respect the BFLPE was similar for boys and girls.

The results have practical implications for educators and coaches in terms of countering the frame-of reference effects associated with the BFLPE. As noted earlier, physical self-concept is important both as an outcome in itself and as a facilitator of short- and long term outcomes. However, the BFLPE has a negative effect on the physical self-concepts of elite athletes. As the BFLPE is based on social comparisons, it is likely to be exacerbated by highly competitive environments and performance feedback that focuses on how athletes compare with other elite athletes. Although there is extensive support for the BFLPE and its unfortunate implications in academic settings, these implications should also generalize to sporting contexts where the focus on competition, social comparison, and winning at all costs is likely to be even stronger than in most academic settings. However, the negative consequences of the BFLPE can, perhaps, be diminished by feedback in relation to self improvement and "personal best" performances. For example, Lüdtke, Köller, Marsh, & Trautwein (2005) found that students' self-concepts were enhanced when teachers evaluating performances used an individualized frame of reference. Beasley (2013) also recommended that physical educators should avoid

> emphasizing personal ability, and providing negative or degrading feedback in front of the class. Instead, teachers can encourage students to focus on personal improvement and task mastery in order to create a positive learning environment that does not highlight social comparison of ability among classmates . . . teachers must be aware of the types of feedback they are providing and strive to give students constructive and corrective feedback regarding skill performance. (p. 35)

Similarly, Liem, Marsh, Martin, McInerney, and Yeung (2013; also see Marsh et al., 2008) argue that feedback should focus on criterion-based assessments and encourage students to pursue personalized and self-referenced goals in relation to previous best performance. Indeed, this is implicitly recognized at the highest level of sport, where personal bests are recognized even in televised broadcasts, along with the winners of gold medals but it has not, perhaps, filtered down to the school level.

There is some support for our speculations about the nature of feedback in the Marsh and Peart (1988) intervention in physical education classes. Students were randomly assigned to one of two experimental groups or a no-treatment control group, in 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 the tasks, the 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 cooperating with a partner. Both experimental interventions significantly enhanced physical fitness relative to pretest scores and in comparison with 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 those participants who performed best on each individual exercise, to a much greater degree than had been the case before 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 initiatives to pursue the 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 our speculations offered here. As emphasized by Trautwein et al. (2008), further research is needed to identify appropriate feedback strategies to enhance physical self-concept and reduce the negative consequences of social comparison and frame-of-reference effects.

Implications for Methodology, Design and Analysis

The present investigation is a substantive-methodological synergy in which we applied the cohort-sequence design with a latent variable approach to address a substantively important issue with theoretical and practical implications. The cohort-sequence design allowed us to draw conclusions across 10 waves spanning a 5-year period, although each participant contributed data in four waves only, spanning 2 of the 5 years. Extending this approach to be fully latent, incorporating multiple indicators of physical self-concept for each of the 10 waves allowed us to test the complex set of invariance assumptions underlying the appropriate interpretation of results based on latent means in general, and the cohort-sequence design in particular, which typically is untested in manifest models. The addition of MIMIC grouping variables (gender, athletic group, and their interaction) allowed us to evaluate stability/change in relation to these groups and test a priori research hypotheses of substantive interest. Although the primary focus of the present investigation is on substantive issues, the latent cohort-sequence approach used here is likely to have broad applicability in sport/exercise research—particularly for research conducted in schools (primary, secondary or university), in which different years in school constitute multiple cohorts.

Limitations and Directions for Further Research

In this study we offer clear a priori predictions based on a well-articulated theoretical model and then offer a new and highly heuristic statistical approach to test these predictions. The change in physical self-concept over high school for athletes and nonathletes might seem paradoxical, in that it might be expected that elite athletes would have substantially higher physical self-concepts than nonathletes and that these differences would increase over time as the athletes matured and improved their sporting expertise relative to those of nonathletes. However, according to the theory underlying the BFLPE, the differences should be highest at the very start of the program (an assimilation or reflected glory effect) but should decline over time as athletes begin to use other athletes as their frame of reference. Although these results are clearly consistent with a priori predictions, a number of limitations dictate that the results should be interpreted cautiously.

In the classic BFLPE paradigm, self-concepts are predicted by individual and group-average performances that are measured in relation to a standardized metric that is common across all participants. Self-concept then is predicted by individual performances and group-average performances (typically the school or class-average achievement in academic studies). Hence, the effect of the group-average variable is negative after controlling for individual performance. In the present investigation, however, there was no common metric with which to evaluate the individual athletic performances of all (elite-athlete and nonathlete) participants. Nevertheless, it is reasonable to assume that athletic accomplishments in the elite-athlete group were substantially higher than those of the nonathletes. Although we cannot establish whether the substantial differences between athletes and nonathletes at the start of high school (M1 in Table 2) are larger than would be expected on the basis of their sporting accomplishments (assimilation), the nonsignificant differences at the end of high school clearly are consistent with a contrast effect. Indeed, if by the end of high school, athletes used only other athletes as their frame of reference, and nonathletes used only other nonathletes, then there should be no differences in the physical self-concepts of the two groups, even though there are substantial differences in the sporting accomplishments of the two groups: this is indeed what we found.

Nevertheless, clear support for a priori hypotheses does not rule out alternative explanations as to the mechanisms underlying the results. This is the case even in studies based on random assignment to experimental and control groups: a significant difference between the groups does not necessarily pinpoint what aspect of the intervention explains it. A useful direction for further research in sport/exercise psychology would be to ask athletes with whom they compare themselves (as in the educational psychology studies of Seaton et al., 2008; Huguet et al., 2009) and to explore the nature of BFLPEs based on qualitative data, asking athletes how they establish frames of reference in forming self-evaluations.

Our study expands support for the claimed universality of the BFLPE as a pan-human theory in relation to academic self-concept (e.g., Seaton, Marsh & Craven, 2009). Although some BFLPE studies have been undertaken in the physical domain, this is apparently the first involving elite athletes. Given that a substantial body of BFLPE research shows the unintended negative consequences of academically selective settings on a wide range of academic outcomes that are mediated at least in part by academic self-concept, and that these negative consequences are not substantially moderated by any of a larger set of potential moderators (e.g., prior ability, learning styles, motivational orientations and goals, background/demographic variables; Marsh, 1991; 2007b; Nagengast & Marsh, 2012; see review by Marsh et al., 2008), it is important to investigate the potential consequences of similar effects in sport/exercise settings. It should be emphasized that the results of the present investigation provide no basis for assuming that the broad-ranging negative implications of selective settings in academic settings generalize to sport/exercise research. Nevertheless, the rich research literature from educational psychology should be heuristic in the further research into the implications of the BFLPE in sport/exercise settings, particularly those involving elite athletes.

Acknowledgments

This article was made possible in part by grants from the Australian Research Council (DP130102713; DP140101559) awarded to the three coauthors.

References

Alicke, M.D., Zell, E., & Bloom, D.L. (2010). Mere categorization and the frog-pond effect. Psychological Science, 21, 174–177. PubMed doi:10.1177/0956797609357718

Beasley, E.K. (2013). Physical self-concept and gender: The role of frame of reference and social comparison among adolescent females (Doctoral dissertation, Mississippi University for Women). Retrieved from http://etd.lsu. edu/docs/available/etd-07082013-100829/unrestricted/ Beasley.Dissertation.Revisions.pdf

Brodbeck, J., Bachmann, M.S., Croudace, T.J., & Brown, A. (2013). Comparing growth trajectories of risk behaviors from late adolescence through young adulthood: An accelerated design. Developmental Psychology, 49, 1732–1738. PubMed doi:10.1037/a0030873

Cohen, P., West, S.G., & Aiken, L.S. (2014). Regression/correlation analysis for the behavioral sciences. New York: Psychology Press.

Dishman, R., O’Connor, P., & Tomporowski, P. (2013). Exercise psychology. Champaign, IL: Human Kinetics.

Dusek, J.B., & Flaherty, J.F. (1981). The development of self-concept during adolescent years. Monographs of the Society for Research in Child Development (Serial No. 191), 46(4), 1–67. Retrieved from http://www.jstor.org

Eccles, J., Wigfield, A., Harold, R.D., & Blumfeld, P. (1993). Age and gender differences in children’s self- and taskperceptions during elementary school. Child Development, 64, 830–847. PubMed doi:10.2307/1131221

Enders, C. (2010). Applied missing data analysis. New York: Guilford Press.

Feingold, A. (1994). Gender differences in personality: A metaanalysis. Psychological Bulletin, 116, 429–456. PubMed doi:10.1037/0033-2909.116.3.429

Fox, K.R., & Corbin, C.B. (1989). The physical self-perception profile: Development and preliminary validation. Journal of Sport & Exercise Psychology, 11(4), 408–430 Retrieved from http://journals.humankinetics.com/jsep.

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

Graham, J.W. (2012). Missing data: Analysis and design. New York: Springer.

Harter, S. (2012). The construction of the self: Developmental and sociocultural foundations. New York: Guilford Press.

Hattie, J. (1992). Self-concept. Hillsdale, NJ: Lawrence Erlbaum Associates.

Huguet, P., Dumas, F., Marsh, H.W., Regner, I., Wheeler, L., Suls, J., . . . & Nezlek, J. (2009). Clarifying the role of social comparison in the big-fish–little-pond effect (BFLPE): An integrative study. Journal of Personality and Social Psychology, 97, 156–170. PubMed doi:10.1037/ a0015558

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

James, W. (1963). The principles of psychology. New York: Holt, Rinehart & Winston

Liem, G.A.D., Marsh, H.W., Martin, A.J., McInerney, D.M., & Yeung, A.S. (2013). The big-fish-little-pond effect and a national policy of within-school ability streaming: Alternative frames of reference. American Educational Research Journal, 50, 326–370. doi:10.3102/0002831212464511

Little, T.D., Slegers, D.W., & Card, N.A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. Structural Equation Modeling, 13(1), 59–72. doi:10.1207/s15328007sem1301_3

Lüdtke, O., Köller, O., Marsh, H.W., & Trautwein, U. (2005). Teacher frame of reference and the big-fish-little-pond effect. Contemporary Educational Psychology, 30, 263–285. doi:10.1016/j.cedpsych.2004.10.002

Luszczynska, A., & Abraham, C. (2012). Reciprocal relationships between three aspects of physical self-concept, vigorous physical activity, and lung function: A longitudinal study among late adolescents. Psychology of Sport and Exercise, 13, 640–648. doi:10.1016/j.psychsport.2012.04.003

Marsh, H.W. (1984). Self-concept: The application of a frame of reference model to explain paradoxical results. Australian Journal of Education, 28, 165–181. Retrieved from http://eric.ed.gov

Marsh, H.W. (1987). The big-fish–little-pond effect on academic self-concept. Journal of Educational Psychology, 79, 280–295. doi:10.1037/0022-0663.79.3.280

Marsh, H.W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early-adulthood. Journal of Educational Psychology, 81(3), 417–430. doi:10.1037/0022-0663.81.3.417

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. doi:10.3102/00028312028002445

Marsh, H.W. (1994). Confirmatory factor analysis models of factorial invariance: A multifaceted approach. Structural Equation Modeling, 1, 5–34. doi:10.1080/10705519409539960

Marsh, H.W. (1996a). Construct validity of Physical Self-Description Questionnaire responses: Relations to external criteria. Journal of Sport & Exercise Psychology, 18(2), 111–131 Retrieved from http://journals.humankinetics.com/jsep.

Marsh, H.W. (1996b). Physical Self-Description Questionnaire: Stability and discriminant validity. Research Quarterly for Exercise and Sport, 67, 249–264. PubMed doi:10.1080/0 2701367.1996.10607952

Marsh, H.W. (1997). The measurement of physical self-concept: A construct validation approach. In K. Fox (Ed.), The physical self: From motivation to well-being (pp. 27–58). Champaign, IL: Human Kinetics.

Marsh, H.W. (1998). Age and gender effects in physical selfconcepts for adolescent elite-athletes and non-athletes: A multi-cohort-multi-occasion design. Journal of Sport & Exercise Psychology, 20(3), 237–259 Retrieved from http://journals.humankinetics.com/jsep.

Marsh, H.W. (2007a). Application of confirmatory factor analysis and structural equation modeling in sport/exercise psychology. In G. Tenenbaum & R.C. Eklund (Eds.), Handbook of sport psychology (3rd ed., pp. 774–798). New York: Wiley.

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

Marsh, H.W., Chanal, J.P., & Sarrazin, P.G. (2006). Self-belief does make a difference: A reciprocal effects model of the causal ordering of physical self-concept and gymnastics performance. Journal of Sports Sciences, 24, 101–111. PubMed doi:10.1080/02640410500130920

Marsh, H.W., & Cheng, J.H.S. (2012). Physical self-concept. In G. Tenenbaum, R.C. Eklund, & A. Kamata (Eds.), Measurement in sport and exercise psychology (pp. 215–226). Champaign, IL: Human Kinetics.

Marsh, H.W., & Craven, R. (2002). The pivotal role of frames of reference in academic self-concept formation: The Big Fish Little Pond Effect. In F. Pajares & T. Urdan (Eds.), Adolescence and Education (Vol. II, pp. 83–123). Charlotte, NC: Information Age Publishing.

Marsh, H.W., Craven, R.G., & Debus, R. (1991). Self–concepts of young children aged 5 to 8: Their measurement and multidimensional structure. Journal of Educational Psychology, 83, 377–392 Retrieved from http://files.eric. ed.gov. doi:10.1037/0022-0663.83.3.377

Marsh, H.W., Craven, R.G., & Debus, R. (1998). Structure, stability, and development of young children's self-concepts: A multicohort-multioccasion study. Child Development, 69(4), 1030–1053. PubMed doi:10.1111/j.1467-8624.1998. tb06159.x

Marsh, H.W., Gerlach, E., Trautwein, U., Lüdtke, U., & Brettschneider, W.D. (2007). Longitudinal study of preadolescent sport self-concept and performance: Reciprocal effects and causal ordering. Child Development, 78, 1640– 1656. PubMed doi:10.1111/j.1467-8624.2007.01094.x

Marsh, H.W., & Hau, K.T. (2007). Applications of latentvariable models in educational psychology: The need for methodological-substantive synergies. Contemporary Educational Psychology, 32, 151–170. doi:10.1016/j. cedpsych.2006.10.008

Marsh, H.W., Hey, J., Roche, L.A., & Perry, C. (1997). The structure of physical self-concept: Elite athletes and physical education students. Journal of Educational Psychology, 89, 369–380. doi:10.1037/0022-0663.89.2.369

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(2), 321–350. doi:10.3102/00028312038002321

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

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

Marsh, H.W., Martin, A.J., & Jackson, S. (2010). Introducing a short version of the Physical Self Description Questionnaire: New strategies, short-form evaluative criteria, and applications of factor analyses. Journal of Sport & Exercise Psychology, 32(4), 438–482. Retrieved from http:// journals.humankinetics.com/jsep. PubMed

Marsh, H.W., & O'Mara, A. (2010). Long-term total negative effects of school-average ability on diverse educational outcomes: Direct and indirect effects of the big-fish-littlepond effect. German Journal of Educational Psychology, 24, 51–72. doi:10.1024/1010-0652/a000004

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

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 & Exercise Psychology, 10(4), 390–407 Retrieved from http://journals.humankinetics.com/jsep.

Marsh, H.W., Perry, C., Horsely, C., & Roche, L. (1995). Multidimensional self-concepts of elite athletes: How do they differ from the general population? Journal of Sport & Exercise Psychology, 17(1), 70–83. Retrieved from http:// journals.humankinetics.com/jsep.

Marsh, H.W., Richards, G.E., Johnson, S., Roche, L., & Tremayne, P. (1994). Physical Self-Description Questionnaire: Psychometric properties and a multitraitmultimethod analysis of relations to existing instruments. Journal of Sport & Exercise Psychology, 16(3), 270–305. Retrieved from http://journals.humankinetics.com/jsep.

Marsh, H.W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K.T., O'Mara, A.J., & Craven, R.G. (2008). The big-fish-little-pondeffect stands up to critical scrutiny: Implications for theory, methodology, and future research. Educational Psychology Review, 20, 319–350. doi:10.1007/s10648-008-9075-6

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

Maxwell, S.E., & Delaney, H.D. (2004). Designing experiments and analyzing data: A model comparison perspective (2nd ed.). Mahwah, NJ: Erlbaum

Nagengast, B., & Marsh, H.W. (2012). Big fish in little ponds aspire more: Mediation and cross-cultural generalizability of school-average ability effects on self-concept and career aspirations in science. Journal of Educational Psychology, 104(4), 1033–1053. doi:10.1037/a0027697

Nesselroade, J.R., & Baltes, P.B. (1979). Longitudinal research in the study of behavior and development. San Diego, CA: Academic Press

Parker, P.D., Marsh, H.W., Morin, A.J.S., Seaton, M., & Van Zanden, B. (2014). If one goes up the other must come down: Examining ipsative relationships between mathematics and English self-concept trajectories across high school. British Journal of Educational Psychology. Advance online publication. PubMed doi:10.1111/bjep.12050

Rubin, D.B. (1987). Multiple imputation for nonresponse in surveys. New York: Wiley.

Schafer, J.L. (1997). Analysis of incomplete multivariate data. New York: Chapman and Hall.

Seaton, M., & Marsh, H.W. (2013). Methodological innovation in big-fish-little-pond effect research. In D.M. McInerney, H.W. Marsh, R.G. Craven, & F. Guay (Eds.), Theory driving research: New wave perspectives on self-processes and human development (Vol. 4, pp. 233–249). Greenwich, Connecticut: Information Age Publishing.

Seaton, M., Marsh, H.W., & Craven, R.G. (2009). Earning its place as a pan-human theory: Universality of the big-fishlittle-pond effect across 41 culturally and economically diverse countries. Journal of Educational Psychology, 101, 403–419. doi:10.1037/a0013838

Seaton, M., Marsh, H.W., Dumas, F., Huguet, P., Monteil, J.-M., Régner, I., . . . Wheeler, L. (2008). In search of the big fish: Investigating the coexistence of the big-fish-littlepond effect with the positive effects of upward comparisons. British Journal of Social Psychology, 47, 73–103. doi:10.1348/014466607X202309

Seligman, M.E.P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. The American Psychologist, 55, 5–14. PubMed doi:10.1037/0003-066X.55.1.5

Shavelson, R.J., Hubner, J.J., & Stanton, G.C. (1976). Validation of construct interpretations. Review of Educational Research, 46, 407–441. doi:10.3102/00346543046003407

Spray, C.M., Warburton, V.E., & Stebbings, J. (2013). Change in physical self-perceptions across the transition to secondary school: Relationships with perceived teacher-emphasised achievement goals in physical education. Psychology of Sport and Exercise, 14(5), 662–669. doi:10.1016/j.psychsport.2013.05.001

Standage, M., Gillison, F.B., Ntoumanis, N., & Treasure, D.C. (2012). Predicting students' physical activity and healthrelated well-being: A prospective cross-domain investigation of motivation across school physical education and exercise settings. Journal of Sport & Exercise Psychology, 34(1), 37–60 Retrieved from http://journals.humankinetics.com/jsep. PubMed

Trautwein, U., Gerlach, E., & Lüdtke, O. (2008). Athletic classmates, physical self-concept, and free-time physical activity: A longitudinal study of frame of reference effects. Journal of Educational Psychology, 100, 988–1001. doi:10.1037/0022-0663.100.4.988

Trautwein, U., Lüdtke, O., Marsh, H.W., & Nagy, G. (2009). Within-school social comparison: How students perceive the standing of their class predicts academic self-concept. Journal of Educational Psychology, 101, 853–866. doi:10.1037/a0016306

Wigfield, A., & Cambria, J. (2010). Students’ achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. Developmental Review, 30(1), 1–35. doi:10.1016/j.dr.2009.12.001

Vallerand, R.J., Blanchard, C., Mageau, G.A., Koestner, R., Ratelle, C., Leonard, M., . . . Marsolais, J. (2003). Les passions de l’ame: On obsessive and harmonious passion. Journal of Personality and Social Psychology, 85, 756–767. PubMed doi:10.1037/0022-3514.85.4.756

Von Eye, A., & Schuster, C. (1998). Regression analysis for social sciences. San Diego: Academic Press

Zell, E., & Alicke, M.D. (2009). Contextual neglect, self-evaluation, and the frog-pond effect. Journal of Personality and Social Psychology, 97, 467–482. PubMed doi:10.1037/ a0015453