Ms 48 11

Promoting a functional physical self-concept in physical education:

Evaluation of a ten-week intervention

Revised manuscript

Author Note

Mirko Schmidt, Institute of Sport Science, University of Bern, Bremgartenstrasse 145, CH-3012 Bern. E-mail: mirko.schmidt@ispw.unibe.ch
Telephone: +41 31 631 83 52 / Fax: +41 31 631 46 31

Stefan Valkanover, Institute of Sport Science, University of Bern, Bremgartenstrasse 145, CH-3012 Bern. E-mail: stefan.valkanover@ispw.unibe.ch
Telephone: +41 31 631 51 56 / Fax: +41 31 631 46 31

Claudia Roebers, Institute of Psychology, University of Bern, Muesmattstr. 45, CH-3009 Bern. E-Mail: roebers@psy.unibe.ch
Tel +41 31 631 47 22 / Fax +41 31 631 82 12

Achim Conzelmann, Institute of Sport Science, University of Bern, Bremgartenstrasse 145, CH-3012 Bern. E-mail: achim.conzelmann@ispw.unibe.ch
Telephone: +41 31 631 83 22 / Fax: +41 31 631 46 31

Correspondence concerning this article should be addressed to Mirko Schmidt, University of Bern, Institute of Sport Science, Bremgartenstrasse 145, CH-3012 Bern.
E-mail: mirko.schmidt@ispw.unibe.ch
Telephone: +41 31 631 83 52 / Fax: +41 31 631 46 31

Abstract

Most physical education intervention studies on the positive effect of sports on self-concept development have attempted to increase school children's self-concept without taking the veridicality of the self-concept into account. The present study investigated whether a 10-week intervention in physical education would lead to an increase not only in the general level of self-concept of endurance and self-concept of strength but also in its veridicality in those who had previously under- or overestimated their abilities. A total of 464 primary school children (246 boys, 218 girls, M_{age} = 11.9) either participated in the intervention or served as controls. The intervention group received an endurance and strength training during physical education lessons carried out with a consistent individualized teacher frame of reference (iTFR). Results showed that this specific intervention was associated with increases not only in the general level of self-concept but also in its veridicality in under- and overestimators. Results are discussed in terms of didactic methods to promote functional self-concepts in physical education.

Key-words: physical self-concept, physical education (PE), veridicality, intervention

Biographical Note

Dr. Mirko Schmidt is a Postdoctoral Research Assistant at the Institute of Sport Science, University of Bern.

Dr. Stefan Valkanover is Senior Lecturer of Sport Pedagogy at the Institute of Sport Science, University of Bern.

Prof. Dr. Claudia Roebers is Senior Professor of Developmental Psychology at the Institute of Psychology, University of Bern.

Prof. Dr. Achim Conzelmann is Senior Professor of Sport Science and Director of the Institute of Sport Science, University of Bern.

Promoting a functional physical self-concept in physical education:

Evaluation of a ten-week intervention

Introduction

Throughout the world, the aims of physical education (PE) are twofold (Liukkonen and Auweele, 2007; Pühse and Gerber, 2005): the first aim is to increase school children's basic motor skills and physical competencies as necessary prerequisites for later participation in physical activities (Weinberg and Gould, 2007). However, the debate on the legitimisation of compulsory PE places even more emphasis on a second aim of PE: its contribution to a positive personality development. The pupils' self-concept (SC) is considered to play a central role in achieving this aim (Bailey, 2006; Vealey, 2002). Given the consistent empirical evidence of a bidirectional relationship between school children's SC and their scholastic achievement (Marsh and O'Mara, 2008), research in the domain of physical activity and PE has focused on the physical SC (Biddle and Mutrie, 2008). As a consequence, most intervention studies in this field have simply tried to increase children's SC, based on the notion that the higher the SC, the more adaptive it is (e.g. Annesi, 2007; Burgess et al., 2006; Calfas and Cooper, 1996; Goñi and Zulaika, 2000; Marsh and Peart, 1988). Besides this assumption of the functionality of a high SC, a second target that has to be focused on from both a pedagogical and a psychological point of view is the veridicality[1] of the SC. Veridicality, which is the reality-relatedness of one's self-perceptions (Helmke, 1992), is

[1] In a broad conceptual understanding, "veridicality" is defined as the extent to which a knowledge structure accurately reflects the information environment it represents (Walsh et al., 1988). Controversy over the accuracy of judgment led to the emergence of a specific research approach addressing the conditions and consequences of accurate (personality) judgments (Funder, 1995). One special case of such accurate judgment is self-judgment or self-appraisal. This transforms the subject into the object being judged. Helmkes' (1992) narrow understanding of "veridicality" (which was used here) as "self-perception appropriate to reality" (p. 197, translated) thereby further narrows the concept from judgment research in order to restrict it to the act of self-perception.

considered to be a functional characteristic in numerous studies from different research areas. For example, it has been reported that an underestimation can have a negative influence on the motivational process (Weiss and Amorose, 2008) and lead to “inadequate task choices” (Harter, 1999), while an overestimation can lead to risky behaviour (Plumert, 1995) or alienation from important social interaction partners (Baumeister et al., 2003). Thus not only the level of the SC but also its veridicality should be considered in interventions promoting a functional SC. The present study used a quasi-experimental approach to positively influence both the level and the veridicality of school children’s physical SC in a 10-week intervention carried out during compulsory PE lessons.

Self-concept: definition and structure

At least since the time of William James (1890), the construct of SC has been an important and much discussed issue in psychology. A number of theorists and researchers have tried to define, describe and differentiate between terms used in various disciplines. In (educational) psychology, there is an abundance of terms in addition to SC such as “self-image”, “self-perceptions”, “self-estimation”, “self-esteem” or “self-worth”, to name just a few. Whereas Bracken and Lamprecht (2003), for example, suggest using the terms SC, self-esteem or self-image synonymously, Dusek and McIntyre (2003) treat these terms differently. From their point of view, SC refers to “the dimensions or categories along which we view the self” and self-esteem refers to “our evaluation or assessment of our self” (p. 290). However, following Marsh and O’Mara (2008), there is, despite minor differences, agreement on the definition of these constructs. Whereas SC, referring to personal characteristics, attributes and abilities, has been seen as the more descriptive, multidimensionally structured component of the self, the term “self-esteem” describes the unidimensional, more evaluative component. Because our focus is on the descriptive component or on the structure of the self, we follow Shavelson, Hubner and Stanton’s (1976) line of argument.

Shavelson et al. (1976) developed a hierarchically organized and multidimensionally structured model of the SC in which the “general self-concept” marks the apex of the pyramid. Moving from the top to the bottom, the term “self-concept” is reserved for judgments in discrete domains such as academic, social, emotional and physical domains, and thus refers to “domain-specific self-evaluations” (Harter, 2005, p. 612). These domains can be further differentiated until one reaches the lowest level containing perceptions of specific competencies or abilities in very specific contexts or situations. We use the terms “facets of self-concept” or “perceived competence” interchangeably to denote this lowest and most specific level of the SC. Whereas the top of the structure is assumed to be relatively stable over time, each level below the self is less stable and more open to change through external forces (Epstein, 1991; Harter, 1999; Shavelson and Bolus, 1982). Numerous studies have supported this specified multi-dimensional structure and shown that the SC cannot be grasped adequately if the aforementioned multidimensionality is not taken into consideration (e.g. Fox and Wilson, 2008; Hagger et al., 2003; Hattie, 1992; Marsh, 1990; Marsh and Craven, 1997; Welk and Eklund, 2005)

The hierarchical structure, the direction of causality and the stability of global and specific dimensions are controversial issues (Kowalski et al., 2003; Marsh and Yeung, 1998) in which at least two models are compared with each other. Bottom-up models assume that changes in global SCs are caused by previous changes in specific dimensions or facets (Harter, 1999; Shavelson et al., 1976). Advocates of top-down models argue that the partial resistance to changes found for specific dimensions or facets can be explained by means of a kind of “buffer function” of the global SC (e.g. Brown, 1993). Although there has been no clear confirmation of the assumed higher stability of global dimensions compared with specific dimensions (Marsh and Hattie, 1996), there is growing evidence to suggest that specific SC dimensions or facets are related more closely to specific behaviour and are more

prone to be influenced by specific interventions than global SCs (Biddle et al., 2000; Marsh and Craven, 2006; O'Mara et al., 2006; Spence et al., 2005).

The veridicality of the self-concept

Two characteristics of SC are crucial in the course of development and in the prediction of an individual's behaviour: first, a positively biased self-perception (i.e. level of SC) is related to many functional characteristics such as functional attributions, (sports or school) performance and job satisfaction (Judge and Bono, 2001; Marsh and Perry, 2005; Valentine et al., 2004; Weiner, 2005). Second, a realistic estimation of one's abilities and skills (i.e. veridicality of SC) influences achievement motivation, adequate task choice and task performance (Försterling and Morgenstern, 2002; Harter, 2006; McFarlin et al., 1984) and may prevent inappropriate risk-taking behaviour (Plumert, 1995; Schwebel and Plumert, 1999).

Weiss and Horn (1990), for instance, have shown how achievement characteristics differ in children with a tendency to over- or underestimate compared to those who assess themselves realistically. Difference scores were calculated on the basis of examinations of their perceived physical competence and their actual competence, and the children were divided into three groups: those who underestimated their skills (underestimators), those who assessed themselves realistically (realists) and those who overestimated themselves (overestimators). Results showed that girls who underestimated their skills had less challenge motivation, higher trait anxiety and higher external control perceptions than girls who assessed themselves realistically. Male underestimators differed from the overestimators and the realists with regard to control perceptions. Phillips and Zimmerman (1990) examined high-achieving students who strongly underestimated their school skills and found that students with this "illusion of incompetence" also had unrealistically low expectations of success and high evaluation anxiety. Moreover, those students rarely chose tasks of adequate difficulty. Harter (1985) has also underlined the problem of inadequate task choice in under- and overestimators. By matching the actual competence of middle school children with

teacher's judgments, she identified three groups: children who overrated their competence, children who underrated their competence and children whose ratings corresponded with those of the teachers. In this study, both overraters and underraters chose easier tasks than the accurate raters. In accordance with Phillips and Zimmerman's (1990) explanation, Harter (1985) argued that underraters were addicted to their illusion of incompetence, whereas overraters preferred less challenging tasks in order to avoid failure and to protect their inflated sense of competence.

Even though the overestimation of one's capabilities is considered to have a number of personal and interpersonal benefits (Taylor and Brown, 1988; Taylor et al., 2003), the limiting factors seem to predominate. Overestimation can lead to "non-productive persistence" (McFarlin et al., 1984), and the overestimation of one's capabilities is often the cause of long-term persistence in pursuing unattainable goals or trying to solve impossible tasks. This leads not only to poorer overall performance (Vancouver et al., 2002) but also to negative consequences for subjective well-being (Carver & Scheier, 1998) or even health (Solberg Nes et al., 2005). In contrast, a realistic assessment of one's capabilities can lead to earlier goal disengagement and thus increase the personal resources – such as time and energy – available to pursue achievable goals (Wrosch et al., 2003).

In social-psychological studies, strong overestimation is associated mostly with negative social consequences (Kim et al., 2010). Overestimators are, for instance, problematic interaction partners because they lack empathy, have high narcissism values and show a higher propensity towards violence (Baumeister et al., 2003; Bushman and Baumeister, 1998; Robins and Beer, 2001). Moreover, although overestimators enjoy high social acceptance in the short term, this shifts in the opposite direction in the long term (Anderson et al., 2006; Colvin et al., 1995; Paulhus, 1998).

The consequences of overestimating abilities can be much more serious in the physical compared to other domains. Plumert (1995), for instance, found that children's overestimation

of their physical abilities was often associated with accidental injuries. Moreover, preschoolers who overestimated their physical abilities, and additionally showed high values on “Extraversion” and low values on “Inhibitory Control”, had more unintentional injuries than all other children (Schwebel and Plumert, 1999). Because assuring children’s safety and health is of major importance not only for parents but also for developmental psychologists and educators, a veridical estimation of one’s own physical abilities has to be one aim of PE and thus also of any interventions in PE settings. Hence, it can be assumed that in a pedagogical and motivational context such as PE, both the level and the veridicality of the perceived competence are of fundamental importance.

Frame-of-reference effects

The literature reveals two major groups of factors that influence an individual’s SC. First, an individual’s performance has both a strong and bidirectional impact (Marsh and Craven, 2006; Marsh and Martin, 2011). Second and going beyond the individual, characteristics of the context in which performance is achieved exert a substantial influence on the level and the veridicality of self-evaluations (Weiss and Amorose, 2008). More specifically, the level of performance within the peer group and the resulting social comparisons (e.g. the big-fish-little-pond effect, BFLPE, Marsh and Hau, 2003) along with, particularly in educational settings, characteristics of the teacher’s behaviour (e.g. teacher-initiated motivational climate) influence children’s self-perceptions, goal orientations and achievement-related cognitions (Ames, 1992; Morgan et al., 2005; Papaioannou et al., 2004).

Empirical support for the BFLPE comes from numerous studies conducted with different samples. Negative effects on SC have been found for students in gifted and talented programs (Marsh et al., 1995; Zeidner and Schleyer, 1999), for students with mild intellectual disabilities (Tracey et al., 2003) as well as for “normal” students (Craven et al., 2000; Lüdtke et al., 2005; Marsh et al., 2007). Although less attention has been paid to the physical SC,

similar findings have been observed in sports (Chanal et al., 2005; Marsh and Perry, 2005) and in PE (Margas et al., 2006).

Although it is impossible (as well as undesirable) to equalize each school child's level of performance when planning interventions in PE, one can target the teacher's behaviour and modify it to fit the specific goals of an intervention. Research has revealed circumscribed educational properties that yield positive effects on the level and veridicality of pupils' SC: avoiding a highly competitive environment that encourages the social comparison processes underlying the BFLPE (Marsh and Peart, 1988), providing students with feedback in relation to criterion reference standards and feedback on their personal improvement over time (Rheinberg and Krug, 2005) and valuing the unique accomplishments of each individual pupil (Marsh and Craven, 2002). It is striking to see how strongly these guidelines for good instruction correspond to a large extent with the guidelines known as "individualized teacher frame of reference" (iTFR, Lüdtke et al., 2005; see related work by Ames, 1992, Covington, 2000) in German-speaking countries.

The concept of TFR was introduced mainly by Rheinberg (1980). Drawing on empirical studies, he noted that teachers differ in their use of reference norms. On the one hand, some compare the performance of a student with those of other students. These teachers use a <u>social reference standard</u>. On the other hand, others assess a student's performance by comparing it with previous performances by the same student. These teachers use an <u>individual reference standard</u>. The central characteristic of instruction based on an iTFR is the way it takes the individual learning progress of each student into consideration when assessing performances. It is assumed that this minimizes the negative effects of social comparison processes, especially for low achieving students, and that this leads to an improvement in the students' SC, interest, motivation and even performance.

Indeed, numerous studies have pointed to the positive effects of the iTFR on the aforementioned constructs (Heckhausen, 1991). Krug and Lecybyl (2005), for instance,

showed that in classes held by teachers with an iTFR, the students' interest increased constantly during the investigation phase, whereas in socially oriented classes, it remained on the same level. Moreover, the assessment of the teacher–student relationship was generally more positive in the iTFR classes. The students' cooperation and their preparation for lessons, as well as participation and performance, especially of the weaker and stronger students, increased continuously towards the end of the intervention.

Sport-specific studies have come to the same conclusions. Krug, Mrazek and Schmidt (1980) modified the teacher's attitude towards an iTFR in order to achieve motivational effects in the students. They conducted a study in PE with fifth-grade pupils. The teacher's encouragement and feedback was based on the iTFR. At the end of the intervention phase, their test group had higher values on hope of success than controls. Krug and Kuhlmann (2005) also examined the motivational effects of the iTFR in PE and found that their fifth-grade pupils gradually set themselves higher goals. In addition, they wanted to exert themselves more often and most of them achieved their goals. The fear of failure decreased significantly during the nine-week intervention phase.

However, attention is focused not only on the demonstrated effects of the iTFR on motivational constructs but also on the direct influence of the iTFR on SC. Since there are no studies on the direct effect of the iTRF on the SC in PE, other school subjects are considered: In their two-year longitudinal study, Rheinberg and Peter (1982) found evidence to suggest that the iTFR has a positive effect on the SC of fifth- to eighth-grade students. In a more recent longitudinal study of 2150 seventh- to eighth-grade students, Lüdtke et al. (2005) found a connection between the iTFR and the students' mathematical SC: The more teachers oriented themselves towards the iTFR, the higher the SCs of their students. This means that in classes with a high iTFR, the students' subject-specific SC develops more favourably than in classes with a low iTFR. The results of the aforementioned studies (cf. Rheinberg & Krug,

2005, for an overview) show that an iTFR can have a positive effect on children's SC and their motivational, affective and achievement-related cognitions.

The present study

Drawing on the different lines of evidence outlined above, the present study evaluated the effects of a 10-week intervention implemented within the context of compulsory PE lessons. The analysis of the possible effects of an intervention was based on the recommendation to use domain-specific SCs, rather than expecting effects in more global SC dimensions (e.g. O'Mara et al., 2006; Spence et al., 2005). This was implemented by defining as targets two specific facets of the SC which are located at the very bottom of the hierarchically structured, multi-dimensional model of the SC (Shavelson et al., 1976), namely the SC of endurance and the SC of strength (Stiller et al., 2004). An intervention was used which considered these two areas and which consisted of ten weeks of systematic endurance and strength training, in order to address the two facets specifically. Due to the fact that not only content has an effect on SC but also teaching style (Marsh and Peart, 1988; Lüdtke et al., 2005), iTFR was used during the entire intervention. This should impede the use of social comparison information and thus have a positive effect on the specific facets of the SC.

The three guiding research hypotheses were the following: A PE intervention which points out the individual learning progress to the students by means of endurance and strength training and an iTFR, will raise the SC of endurance and the SC of strength (Hypothesis 1). The intervention should not only have an effect on the level of the SC of the facets addressed, but also on their veridicality. It is thus assumed that after the intervention, the students will be able to assess themselves more accurately than before the intervention. The respective differential hypothesis is the following: While the veridicality of the overestimators' and underestimators' SC of endurance and SC of strength increases, it remains on the same level for the realists (Hypothesis 2). Based on a bottom-up model of SC (Shavelson et al., 1976), which postulates that global SCs are more stable over time than domain-specific SC

dimensions, it is assumed that the same intervention will have no effects on the general SC (Hypothesis 3).

Method

Design

The present paper deals with part of the question raised in BISS (the Bernese Intervention Study on School Sport), a quasi-experimental longitudinal study aimed at analysing the effects of PE lessons on the SC of fifth grade students (Conzelmann, Schmidt and Valkanover, 2011). Three interventions were developed, which were based on changes in the three non-academic SC dimensions (Shavelson et al., 1976). While the intervention “play” is expected to influence the social SC positively, the intervention “venture” should have a positive effect on the emotional, and the intervention “performance” on the physical SC. Over the course of half a school year, two ten-week treatment phases were conducted. Each intervention phase was preceded and followed by a measurement point (T1, T2, T3) for data collection (classical comparison group pre-test/post-test design). 17 intervention classes went through the treatment phases in a different order. With three different treatment programmes, there are six possible permutations (performance-play, performance-venture, play-performance, play-venture, venture-performance, venture-play), each of which was used in two or three classes. The control group (6 classes) had normal PE lessons without psychological treatment during the entire intervention phase. Each of the measurements included a questionnaire survey of the SC and sports motor tests. This paper deals with the intervention “performance”, which is expected to have a positive effect on the SC of endurance and SC of strength.

Sample

A total of 464 primary school children (47% female) from 23 schools in urban and rural areas of Switzerland participated in the study. The intervention group (n = 246; 126 boys, 120 girls;

M_{age} = 11.9, SD = .56) includes those students who underwent the performance intervention either in the first or the second treatment phase. Mixing participants who completed the treatment program in different orders seems acceptable since no carry-over effects (Lewis-Beck et al., 2004) are observed: The difference in the difference scores (M_{T2-T1} minus M_{T3-T2}) between these two groups is not significant either for the SC of endurance ($t(244) = -.26$, $p = .796$, $d = .001$) or for the SC of strength ($t(244) = -.54$, $p = .589$, $d = .066$). The control group ($n = 218$; 120 boys, 98 girls; M_{age} = 12.0, SD = .53) includes those students who underwent a different or no psychological treatment programme. The two groups do not differ with regard to age ($\chi^2 = 4.05$, d.f. = 3, $p = .256$, Cramer's $V = .11$), sex ($\chi^2 = .59$, d.f. = 1, $p = .444$, Cramer's $V = .04$) or social background ("Family Affluence Scale", Currie et al., 1997; $\chi^2 = 5.59$, d.f. = 8, $p = .693$, Cramer's $V = .14$). There was some loss of data due to sick leave, non-participation in the sports motor tests because of injury or incomplete questionnaires. The percentage of pupils with incomplete values was 4.7% at pre-test and 8.8% at post-test. The resulting missing values were augmented with the help of the method of multiple imputation - which is briefly described in the Data Analysis section - so that it was possible to work with a complete set of data.

The teacher sample consisted of 23 individuals, with a higher proportion of women (60.9 %) than men. This approximately matches the unequal gender distribution in pedagogic professions. The teachers' mean age at the first measurement was M = 33.83 years (SD = 9.02). Their professional experience ranged from 1.5 to 29 years with a mean of M = 11.02 years (SD = 7.94). The two groups do not differ with regard to sex ($\chi^2 = .35$, d.f. = 1, $p = .552$, Cramer's $V = .12$), age ($\chi^2 = 15.99$, d.f. = 16, $p = .454$, Cramer's $V = .83$) or professional experience ($\chi^2 = 14.65$, d.f. = 14, $p = .402$, Cramer's $V = .79$).

Measures

Perceived physical competence

The physical SC was measured using the two 6-item subscales “endurance” and “strength” from the Physische Selbstkonzept-Skalen [Physical self-concept scales] (PSK; Stiller et al., 2004). This instrument was derived from a former version of the Physical Self-Description Questionnaire (PSDQ; Marsh et al., 1994). An example of an endurance item is “I can run a long way without stopping”. An example of a strength item is “I would do well in a test that measures strength”. All items were rated on a 4-point scale ranging from 1 (strongly disagree) to 4 (strongly agree). Cronbach’s alpha ranged from .85 to .88, thus showing high internal consistency.

Actual physical competence

The children’s endurance performance was assessed with the Multistage 20 Meter Shuttle Run Test (Léger et al., 1988). Subjects have to run back and forth on a 20 m course and touch the 20 m line with their foot, and at the same time, a sound signal is emitted from a pre-recorded tape. The frequency of the sound signal increases by 0.5 km/h every minute, starting with a speed of 8.5 km/h. The test ends when subjects fail to reach the line before the signal twice in a row. The test item score is the time achieved in seconds. The test-retest reliability coefficient was .80 in our sample, as compared with .89 for boys and girls aged 8–19 years (Léger et al., 1988). Further evidence on its reliability and validity has been provided by Liu et al. (1992) and McVeigh et al. (1995).

The actual maximal strength of the upper extremities was assessed using a shot-put performance test. While sitting on a bench with a height of 38 cm, the pupils had to push a medicine ball (weighing 2 kg) as far as possible using both hands (Bös and Wohlmann, 1987). Their feet were parallel to each other, their shoulders were pressed against a wall and they held the ball in front of their chest. The test item score (best of two tries) was the distance

achieved in meters. Validity and reliability have already been demonstrated (Wyss et al., 2007) and the test-retest reliability coefficient was .81 in our sample.

Veridicality of perceived physical competence

In order to form the veridicality variable, it is first necessary to know the students' main reference group. Whether students compare themselves mainly with all students of the same age, with students of the same sex or with students in the same class can, according to Marsh (1993), be assessed by means of three correlation calculations. Therefore the two variables needed to depict veridicality must be z-standardized to (a) the class, (b) the specific sex and (c) the total sample. It can then be assumed that the students orient themselves towards the reference group that proves to have the highest correlation values. Results of these calculations showed that participants evaluated their performance of endurance (strength in parenthesis) most strongly within their class (r = .59 (.34), .48 (.26), and .54 (.31), for the class, same-sex peers, and the entire sample, respectively). Therefore, all variables were z-standardized by class. Veridicality was determined by calculating difference scores between the SC values (perceived endurance) and actual endurance. To analyse the differential hypothesis, children were divided into three groups which differed in terms of the veridicality of their SC: those with values of $z < -1$ ($n_{Endurance}$ = 64, $n_{Strength}$ = 87) were labelled underestimators; those with values of z = -1 to 1 ($n_{Endurance}$ = 341, $n_{Strength}$ = 292), as realists; and those with values of $z > 1$ ($n_{Endurance}$ = 59, $n_{Strength}$ = 85), as overestimators.

General self-concept

In order to measure general self-concept, the German version of the Rosenberg Self-Esteem Scale (Rosenberg, 1965) was used (von Collani and Herzberg, 2003). The short form of the scale consists of 5 items, one example of which is: "I feel that I have a number of good qualities". All items were rated on a 4-point scale ranging from 1 (strongly disagree) to 4 (strongly agree). Cronbach's alpha ranged from .74 to .84.

Individualized teacher frame of reference

Intervention integrity (i.e., the degree to which interventions are implemented as planned; Dane and Schneider, 1998) was checked through students' perceptions of the teachers' frame of reference (Schwarzer and Jerusalem, 1999). This adapted scale was used to ask pupils to what extent their teacher applied an iTFR in her or his PE instruction. The short version used here contained three items such as "Our PE teacher particularly praises pupils when they perform better than they did before". All items were rated on a 4-point scale ranging from 1 (strongly disagree) to 4 (strongly agree). Cronbach's alpha ranged from .73 to .77.

Procedure

The study was carried out in cooperation with schools in or around the city of Bern, Switzerland. The first step was to inform the canton and city authorities about our research plans and obtain formal permission to approach school principals. The second step was to write to all school principals in and around the city informing them about the goals of the project, the assessment methods and the time plan. After receiving their principals' permission, 23 interested 5th-grade teachers were contacted, who agreed to commit themselves to participating in the project.

Even though it was not possible to completely exclude, hold constant and randomize confounding variables (as is the standard in laboratory experiments), participants were matched on known socio-economic confounds (social background, urban vs. rural) in order to form homogeneous research groups.

Pre-test data were collected during two consecutive school lessons. During the first lesson, the children completed the questionnaire; in the second lesson, they carried out the endurance and strength tests in the gym. The same consecutive procedure was also used for the post-test after the intervention. Parents gave their informed consent prior to the study.

Intervention

The intervention was carried out by the teachers of the intervention group in two out of the three compulsory PE lessons per week over a period of 10 weeks. Hence, the entire intervention extended over twenty 45-minute lessons. Teachers previously completed a half-day training programme instructing them in the basic principles, aims and purposes of the intervention programme and demonstrating the specific contents with special teaching materials. In order to increase the chances of implementing the strategies learned during training, the following framework conditions were implemented on the basis of previous research findings (McCoy and Reynolds, 1998; Paul and Volk, 2002; Yamnill and McLean, 2001): Firstly, the content taught during teacher training was based on existing teaching material that was well known to the teachers. Secondly, concrete contents were demonstrated directly on the basis of teaching material which was handed out to the teachers after the training. Thirdly, the theory was of strong practical relevance and based on case studies. Finally, the teachers were supported on the phone during the entire intervention period in order to eliminate any uncertainties that arose. The phone calls took place twice during each intervention period. The questions raised by the teachers were primarily ones of comprehension concerning the teaching materials supplied to them, all of which were satisfactorily answered.

The intervention programme was characterized by two features, namely content and method. Endurance and strength were chosen because other studies have shown that the training of these two components had the biggest effect on the specific facets of the SC (e.g. Biddle et al., 2000; Spence et al., 2005). While the endurance training aimed at improving participants' running performance, the strength training mainly focused on the upper extremities. To guarantee an age-appropriate training, playful forms of running exercises (with music) and strength training were used to improve the endurance and the strength performance. With regard to method, the intervention was based on the iTFR, with teachers

receiving didactic method guidelines focusing on five central sources for children's positive self-evaluations: (a) making children aware of their own behaviour and performance (self-perceptions) by documenting performance progress explicitly in a specially designed sports booklet, by asking questions that encourage self-reflection and by providing opportunities for self-observation (Bem, 1972); (b) giving children direct, frequent and positive feedback (e.g. by means of positive feedback during motivational challenges) (Mouratidis et al., 2010); (c) establishing a fear-free atmosphere to increase the children's mutual social acceptance and reduce indirect negative feedback from peers; (d) giving the children the chance to pursue self-chosen goals, thereby making it possible to collect the ideational self-information that forms the critical basis for the actual self-image (e.g. "possible selves"; Markus and Nurius, 1986); and (e) avoiding upward social comparisons when giving personal feedback (Suls et al., 2002) while accepting that social comparison processes cannot be avoided or inhibited completely. This complete set of didactic methods was implemented in order to minimize the BFLPE and to maximize the children's perception of their personal improvements over time.

Teachers in the control group continued to teach according to the national curriculum for PE (Federal Agency of Sport, 1997). Every week, they had to document the contents and goals of their teaching in a table. These tables showed that they carried out their usual PE without any major changes.

Data analyses

All the analyses were carried out using the program IBM SPSS Statistics 19.

In order to determine the effect of the intervention on the pupils' SC, methods for the analysis of variance (with repeated measurement) were used, as is usual in such quasi-experimental designs. Since the hypotheses were worded directionally, a one-sided test was used. A significance level of .05 was used for all tests. In order to estimate the effect size, $\eta^2_{time} = (SS_{time})/(SS_{time} + SS_{total})$, $\eta^2_{group} = (SS_{group})/(SS_{group} + SS_{total})$, and $\eta^2_{time\ x\ group} = (SS_{time\ x\ group})/(SS_{time\ x\ group} + SS_{total})$ were quoted in the ANOVAs, whereby $\eta^2 = .0099$ represents

a small effect, $\eta^2 = .0588$ a moderate and $\eta^2 = .1379$ a large effect (Cohen, 1988). For the t-tests, Cohen's d = (mean difference)/(standard deviation) was calculated, whereby d = .2 represents a small, d = .5 represents a moderate and d = .8 represents a large effect (Cohen, 1992).

Missing values can lead to unwanted distortions in statistical analyses. Our aim was therefore to remove missing values. Two different approaches are generally available for this: case deletion or imputation procedures (Graham, 2009). Case deletion procedures are only permissible, (a) if there are few missing values (roughly speaking less than 5 % of the overall number of cases) and (b) if the values can be considered to be missing completely at random (MCAR; Schafer and Graham, 2002). In the present study, 8.8% of cases displayed missing values at the post test, and the MCAR test according to Little proved significant ($\chi^2 = 530.38$, d.f. = 465, p = .019), so the missing values were imputed using multiple imputation (MI). The process of MI comprises three separate steps (Rubin, 1987). In the first step, several replacements are carried out for each missing value, taking into account the information available from the data set, leading to complete sets of data. In the second step, each of these complete sets of data is then analysed using standard methods (e.g. ANOVA). In the final step, the results of the separately conducted analyses are combined, taking into account the uncertainty of the imputation process. This means that in MI the imputation model (first step) is kept separate from the analysis model (second step) and the statistical inference (third step) is carried out separately from the statistical analysis. In the present study, m = 10 imputations were carried out, in line with Acock's recommendation (2004). The maximum number of iterations was set at k = 10, so that for each imputation the Markov Chain Monte Carlo (MCMC) algorithm was performed ten times. In addition, all variables were defined both as predictors and as imputation variables in the complete conditional specification.

Results

Preliminary analyses

Table 1 presents the descriptive statistics for the study variables and the results of the preliminary analyses. All dependent variables were normally distributed with skewness values of -.09 to .58 and kurtosis values of -1.09 to 1.55. Independent t tests showed that none of the dependent variables differed between the intervention and the control group at pre-test (ps > .20). All dependent variables except the iTFR revealed significant gender differences with small to medium effect sizes (ds = .26 – .64). Because the inclusion of gender as an additional factor did not lead to significant interaction effects, it is not reported here.

[Table 1]

Whether teachers in the intervention group really implemented the intervention programme was tested by means of a mixed ANOVA in which the students' perception of their teachers' frames of reference (individualized TFR) was treated as a dependent variable, with the group (intervention vs. control) as the between-subject factor and time (pre- vs. post-test) as the within-subject factor. There was no main effect of either time, $F(1, 462) = 2.00$, $p = .157$, $\eta^2 = .004$, or group, $F(1, 462) = .43$, $p = .513$, $\eta^2 = .001$, but a significant interaction between the two, $F(1, 462) = 2.76$, $p = .048$, $\eta^2 = .006$, suggesting that teachers in the intervention group used an increased iTFR during the intervention time.

Before testing the three main hypotheses, a mixed ANOVA was conducted with the group (intervention vs. control) as the between-subject factor and time (pre- vs. post-test) as the within-subject factor, to test whether the intervention was associated with an increase in the actual endurance and the actual strength. The left side of Figure 1 shows that actual endurance increased to the same extent in both the intervention and the control group. There was a significant main effect of time, $F(1, 462) = 91.35$, $p < .0005$, $\eta^2 = .165$, but not of group, $F(1, 462) = .16$, $p = .689$, $\eta^2 = .000$. Because there was no significant interaction

between time and group, $F(1, 462) = .04$, $p = .834$, $\eta^2 = .000$, it can be concluded that the intervention had no specific impact on children's actual endurance. The right-hand side of Figure 1 shows a similar but moderate pattern for the actual strength: There was a significant main effect of time, $F(1, 462) = 5.60$, $p = .018$, $\eta^2 = .012$, but not of group, $F(1, 462) = .25$, $p = .619$, $\eta^2 = .001$. With no significant interaction between time and group, $F(1, 462) = .42$, $p = .519$, $\eta^2 = .001$, it could be deduced that the intervention had no effect on children's actual strength either.

[Figure 1]

Primary analyses

The next step in the analyses explored whether there was a specific effect on the SC of endurance and the SC of strength (Hypothesis 1). While the SC of endurance and the SC of strength respectively were treated as dependent variables, the group (intervention vs. control) was treated as a between-subject factor and time (pre vs. post) as a within-subject factor. The left side of Figure 2 shows that the intervention was associated with an increase in the SC of endurance. There was a significant main effect of time, $F(1, 462) = 9.16$, $p = .003$, $\eta^2 = .019$, but not of the group, $F(1, 462) = .23$, $p = .629$, $\eta^2 = .001$. Importantly, there was a significant interaction between time and group: the SC of endurance increased more strongly in the intervention group than in the control group, $F(1, 462) = 14.38$, $p < .0005$, $\eta^2 = .030$, suggesting a specific effect of the intervention on the SC of endurance. With regard to the SC of strength (right-hand side of Figure 2), the main effects are similar. The ANOVA yields a significant main effect of time $F(1, 462) = 5.99$, $p = .015$, $\eta^2 = .013$, but not of group, $F(1, 462) = .00$, $p = .969$, $\eta^2 = .000$. In contrast to the SC of endurance, no significant interaction is observed for the SC of strength $F(1, 462) = .70$, $p = .403$, $\eta^2 = .002$. Contrary to the expectations, the intervention had no positive effect on the SC of strength in the intervention group that could not also be found in the control group.

[Figure 2]

To test Hypothesis 2, an ANOVA with repeated measures was again conducted, but the veridicality of the SC of endurance and the veridicality of the SC of strength respectively served as dependent variables. Since the second hypothesis is a differential hypothesis, the three groups of overestimators, realists and underestimators were analysed separately.

Table 2 shows that the two groups of over- and underestimators exhibited main effects of time, but not of group. More importantly, and as hypothesized, there were significant interactions between time and group. Figure 3 shows that the veridicality of the SC of endurance and the veridicality of the SC of strength in the intervention group improved for both overestimators and underestimators, compared with controls. There were no significant main or interaction effects of the intervention on the veridicality of the SC of endurance and on the veridicality of the SC of strength in the realists (all ps > .215).

[Figure 3]

[Table 2]

In order to test whether a specific intervention only has an effect on the desired SC dimensions and not on other SC dimensions as well (Hypothesis 3), the effect of the intervention on the dimension of general SC was checked, which is higher up in the hierarchy. With regard to general SC, the ANOVA revealed no significant main effect of time, $F(1, 462) = .03$, $p = .866$, $\eta^2 = .000$, nor of the group, $F(1, 462) = 1.24$, $p = .266$, $\eta^2 = .004$. Because there was no significant interaction between time and group, $F(1, 462) = .08$, $p = .784$, $\eta^2 = .000$, it can be concluded that the intervention had no specific impact on children's general SC.

In conclusion, it can be stated that the ten-week intervention, which was specifically based on the SC of endurance and the SC of strength, had a positive effect on the level of the SC of endurance, but not on the level of the SC of strength. Moreover, the overestimators and underestimators of the intervention group assessed their endurance and strength with greater veridicality after the intervention than the overestimators and underestimators in the control

group, while the realists' veridicality of the respective SC facets was not affected by the intervention. Because there was no effect on the general SC, the intervention seems to have some positive effects on intended and domain-specific SC facets, but no effects on non-intended, more global dimensions of the SC.

Discussion

The results of this study show that a specific intervention in PE may be associated with an increase in the level of school children's SC of endurance. Whereas previous interventions have been implemented in after-school programmes (Annesi, 2007; Marsh and Peart, 1988) or have been stretched out over much longer periods of time (e.g. Goñi and Zulaika, 2000), the present intervention in PE was associated with positive effects on specific SC facets in only 10 weeks. As the measurement of performance showed, school children's endurance and strength can be increased within PE (see Figure 1); however, the additional aim of fostering a positive SC was achieved only through endurance training in combination with specific and focused didactic methods. The fact that the desired effect was not observed for the SC of strength might be explained by the intensity and the specificity of the lessons during the intervention. In the field of endurance, the weekly lesson was used to improve running performance, whereas the weekly lesson to improve strength was further divided into different parts. Only about half of the lesson was devoted to appropriate strengthening exercises. This may also explain the small degree of change in the actual strength. Of course, the present study is limited to positive effects on one SC facet, i.e. the SC of endurance, and it remains to be shown whether such effects can also be obtained for other SC facets. Hence, the present approach offers an encouraging starting point for future research. Nevertheless, within the broader discussion of sustained effects on SC development and their transfer to everyday activities, the important question of whether increases in multiple SC facets can lead to long-term enhancements in higher level (e.g., global) dimensions of SC still remains unanswered.

The present approach appears to be the first attempt to increase the veridicality of school children's SC. This is rather surprising in the light of the wealth of studies on the functionality of realistic self-assessment (e.g. Försterling and Morgenstern, 2002; McFarlin et al., 1984; Phillips and Zimmerman, 1990; Schwebel and Plumert, 1999; Weiss and Horn, 1990). The study reported here shows clearly that children who either strongly over- or underestimate their performance assess their endurance and strength more realistically after the intervention, compared with student controls. This is a notable finding, especially since there were moderate to large effects suggesting that meaningful changes had occurred. In an educational context, a veridical SC is of great importance, because it is associated with higher achievement motivation (Weiss and Amorose, 2008), adequate task choice (Harter, 1999) or decreased risk-taking behaviour (Plumert, 1995). Moreover, the present results document a means to promote this.

Turning to the broader discussion about the hierarchical structure, the direction of causality and the stability of global and specific SC dimensions and facets (Kowalski et al., 2003; Marsh and Yeung, 1998), the present study makes a contribution to empirically confirming the postulated multi-dimensional and hierarchical structure of SC (Shavelson et al., 1976). Whereas the intervention was associated with small to medium effects on the lowest SC facets, it had no influence on the more global construct of the general SC at the top of the hierarchy. Global SC dimensions seem to be more stable than specific SC facets. Hence, the findings can be explained more easily on the basis of a bottom-up (Harter, 1999; Shavelson et al., 1976) rather than a top-down model (Brown, 1993). It can be concluded that future research on the relationship between sports and SC development should focus on area-specific SC facets in shorter interventions and on global SC dimensions in longer interventions. O'Mara et al. (2006) also come to this conclusion and showed that it was mainly those interventions that aimed to change the lower dimensions or facets of the SC rather than general SC or self-esteem that had an effect on their participants. It can therefore

be said that specifically designed interventions are needed to address specific SC facets. The more the selected contents and methods correspond to the facets, the more likely they are to be changed.

The intervention was associated with a positive influence on both the level and the veridicality of the children's SC of endurance and the veridicality of the children's SC of strength. These results are not just attributable to the choice of contents, but also to their didactic implementation. It has been suggested that the individualized teacher frame of reference and the resulting feedback have demonstrably positive consequences for children's SC and other cognitive personality variables (Krug et al., 1980; Lüdtke et al., 2005; Rheinberg and Peter, 1982). The present findings seem to confirm these assumptions, because increases in actual endurance were obtained both with and without the described specific intervention focusing on an individualized teacher frame of reference. However, alongside the special content of the intervention, the central difference between the two groups of students was the aspect of conscious consideration in the sense of a guided reflection on their own performance and the improvement in their own performance. Therefore, the findings can be interpreted as showing that the significant effects of this study are due to the consistently supported reflection on one's own performance in the form of a feedback dialogue between children and teachers or between children and their peers, along with the recording of one's individual learning development and assessment in the sports booklet.

With regard to the teachers, the implementation methods applied in the current intervention require distinctly individualized PE accompanied by context manipulations designed to systematically decrease social comparisons between classmates. The present approach transferred learning methods from other school subjects into the gym by means of discourse and paperwork. One could argue that this limits its feasibility, because its didactic intervention methods seem to intensify preparation while limiting the compensatory function of PE in the normal school day. Nonetheless, our findings do show that selectively including

cognitive methods in PE lessons is an effective way of achieving the goal of developing functional SC embedded in both the curriculum and educational policy.

Acknowledgement

The study is part of a larger project entitled ‘The Bernese Intervention Study on School Sports (BISS)’ supported by a grant from the Federal Agency of Sport, Switzerland.

References

Acock, A. C. (2005) ‘Working with Missing Values‘, Journal of Marriage and Family 67: 1012-1028.

Ames, C. (1992) ‘Classroom Goals, Structures, and Student Motivation’, Journal of Educational Psychology 84: 261-271.

Anderson, C., Srivastava, S., Beer, J. S., Spataro, S. E. and Chatman, J. A. (2006) ‘Knowing Your Place: Self-Perceptions of Status in Face-to-Face Groups’, Journal of Personality and Social Psychology 91: 1094-1110.

Annesi, J. J. (2007) ‘Relations of Age with Changes in Self-Efficacy and Physical Self-Concept in Preadolescents Participating in a Physical Activity Intervention During Afterschool Care’, Perceptual and Motor Skills 105: 221-226.

Bailey, R. (2006) ‘Physical Education and Sport in Schools: A Review of Benefits and Outcomes’, Journal of School Health 76: 397-401.

Baumeister, R. F., Campbell, J. D., Krueger, J. I. and Vohs, K. D. (2003) ‘Does High Self-Esteem Cause Better Performance, Interpersonal Success, Happiness, or Healthier Lifestyles?’ Psychological Science in the Public Interest 4: 1-44.

Bem, D. J. (1972) ‘Self-Perception Theory’, in L. Berkowitz (ed) Advances in Experimental Social Psychology, pp. 1-62. New York: Academic Press.

Biddle, S. J. H., Fox, K. R. and Boutcher, S. H. (2000) Physical Activity and Psychological Well-Being. London: Routledge.

Biddle, S. J. H. and Mutrie, N. (2008) Psychology of Physical Activity: Determinants, Well-Being and Interventions, 2nd edn. London: Routledge.

Bös, K. and Wohlmann, R. (1987) 'Allgemeiner Sportmotorischer Test (AST 6-11) Zur Diagnose der konditionellen und koordinativen Leistungsfähigkeit'[Diagnosis of Conditional and Coordinative Performance], Lehrhilfen für den Sportunterricht 36: 145-160.

Bracken, B. A. and Lamprecht, M. S. (2003) 'Positive Self-Concept: An Equal Opportunity Construct', School Psychology Quarterly 18: 103-121.

Brown, J. D. (1993) 'Self-Esteem and Self-Evaluation: Feeling Is Believing', in J. M. Suls (ed.) Psychological Perspectives on the Self, pp. 27-58. Hillsdale, NJ: Erlbaum.

Burgess, G., Grogan, S. and Burwitz, L. (2006) 'Effects of a 6-Week Aerobic Dance Intervention on Body Image and Physical Self-Perceptions in Adolescent Girls', Body Image 3: 57-66.

Bushman, B. and Baumeister, R. F. (1998) 'Threatened Egotism, Narcissism, Self-Esteem, and Direct and Displaced Aggression: Does Self-Love or Self-Hate Lead to Violence?' Journal of Personality and Social Psychology 75: 219-229.

Calfas, K. and Cooper, D. (1996) 'Effect of a 5 Week Exercise Training Program on Self-Worth among Adolescent Girls: A Randomised Controlled Study', Medicine and Science in Sports and Exercise 28: 135.

Carver, C. S. and Scheier, M. F. (1998) On the Self-Regulation of Behavior. New York: Cambridge University Press.

Chanal, J. P., Marsh, H. W., Sarrazin, P. G. and Bois, J. E. (2005) 'Big-Fish-Little-Pond Effects on Gymnastics Self-Concept: Social Comparison Processes in a Physical Setting', Journal of Sport and Exercise Psychology 27: 53-70.

Cohen, J. (1988) Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Erlbaum.

Cohen, J. (1992) 'A Power Primer', Psychological Bulletin 112: 155-159.

Colvin, C. R., Block, J. and Funder, D. C. (1995) 'Overly Positive Self-Evaluations and Personality: Negative Implications for Mental Health', Journal of Personality and Social Psychology 68: 1152-1162.

Conzelmann, A., Schmidt, M. and Valkanover, S. (2011) 'Persönlichkeitsentwicklung durch Schulsport. Theorie, Empirie und Praxisbausteine der Berner Interventionsstudie Schulsport (BISS)' [Personality Development through Physical Education. Theory, Empirical Evidence, and Practical Implications of the Bernese Intervention Study on School Sports (BISS)]. Bern: Huber.

Covington, M. V. (2000) 'Goal Theory, Motivation, and School Achievement: An Integrative Review', Annual Review of Psychology 51: 171-200.

Craven, R. G., Marsh, H. W. and Print, M. (2000) 'Selective, Streamed and Mixed-Ability Programs for Gifted Students: Impact on Self-Concept, Motivation and Achievement', Australian Journal of Education 44: 51-75.

Currie, C. E., Elton, R. A., Todd, J. and Platt, S. (1997) 'Indicators of Socio-Economic Status for Adolescents: The WHO Health Behaviour in School-Aged Survey', Health Education Research 12: 385-397.

Dane, A. V. and Schneider, B. H. (1998) 'Program Integrity in Primary and Early Secondary Prevention: Are Implementation Effects out of Control?' Clinical Psychology Review 18: 23-45.

Dusek, J. B. and McIntyre, J. G. (2003) 'Self-Concept and Self-Esteem Development', in G. R. Adams and M. D. Berzonsky (eds) Blackwell Handbook of Adolescence, pp. 290-309. Malden, MA: Blackwell.

Epstein, S. (1991) 'Cognitive-Experiential Self-Theory: An Integrative Theory of Personality', in R. Curtis (ed.) The Self with Others: Convergences in Psychoanalytic, Social, and Personality Psychology, pp. 111-137. New York: Guilford Press.

Federal Agency of Sport (1997) Lehrmittel Sporterziehung Band 4 [Teaching aids for physical education, vol. 4]. Bern: EDMZ.

Försterling, F. and Morgenstern, M. (2002) 'Accuracy of Self-Assessment and Task Performance: Does It Pay to Know the Truth?' Journal of Educational Psychology 94: 576-585.

Fox, K. R. and Wilson, P. M. (2008) 'Self-Perceptual Systems and Physical Activity', in T. S. Horn (ed.). Advances in Sport Psychology, 3rd ed., pp. 49-64. Champaign, IL: Human Kinetics.

Funder, D. C. (1995) 'On the Accuracy of Personality Judgment: A Realistic Approach', Psychological Review 102: 652-670.

Goñi, A. and Zulaika, L. (2000) 'Relationships between Physical Education Classes and the Enhancement of Fifth Grade Pupils' Self-Concept', Perceptual and Motor Skills 91: 246-250.

Graham, J. W. (2009) 'Missing Data Analysis: Making It Work in the Real World', Annual Review of Psychology 60: 549-576.

Hagger, M. S., Biddle, S. J. H., Chow, E. W., Stambulova, N. and Kavussanu, M. (2003) 'Physical Self-Perceptions in Adolescence: Generalizability of a Hierarchical Multidimensional Model across Three Cultures', Journal of Cross-Cultural Psychology 34: 611-628.

Harter, S. (1985) 'Competence as a Dimension of Self-Evaluation: Toward a Comprehensive Model of Self-Worth', in R. L. Leahy (ed.) The Development of the Self, pp. 55-121. New York: Academic Press.

Harter, S. (1999) ‘The Construction of the Self: A Developmental Perspective‘. New York: Guilford Press.

Harter, S. (2005) ‘The Development of Self-Representations during Childhood and Adolescence‘, in M. R. Leary & J. P. Tangney (eds), Handbook of Self and Identity (6th ed., pp. 610-642). New York, NY: Guilford.

Harter, S. (2006) ‘The Self‘, in W. Damon & R. M. Lerner (eds), Handbook of Child Psychology (Vol. 6). New York, NY: John Wiley & Sons.

Hattie, J. (1992) Self-Concept. Hillsdale, NJ: Erlbaum.

Heckhausen, H. (1991) Motivation and Action. New York: Springer.

Helmke, A. (1992) Selbstvertrauen und schulische Leistungen [Self-Confidence and School Performance]. Göttingen: Hogrefe.

James, W. (1890) The Principles of Psychology. New York: Holt.

Judge, T. A. and Bono, J. E. (2001) ‘Relationship of Core Self-Evaluations Traits - Self-Esteem, Generalized Self-Efficacy, Locus of Control, and Emotional Stability - with Job Satisfaction and Job Performance: A Meta-Analysis’, Journal of Applied Psychology 86: 80-92.

Kim, Y.-H., Chiu, C.-y. and Zou, Z. (2010) ‘Know Thyself: Misperceptions of Actual Performance Undermine Achievement Motivation, Future Performance, and Subjective Well-Being’, Journal of Personality and Social Psychology 99: 395-409.

Kowalski, K. C., Crocker, P. R. E., Kowalski, N. P., Chad, E. and Humbert, M. L. (2003) ‘Examining the Physical Self in Adolescent Girls over Time: Further Evidence against the Hierarchical Model’, Journal of Sport & Exercise Psychology 25: 5-18.

Krug, S. and Kuhlmann, K. (2005) ‘Motiveffekte individueller Bezugsnormen im Sportunterricht’ [Motivational effects of individual reference norms in physical education], in F. Rheinberg (ed.) Motivationsförderung im Schulalltag [Promoting motivation in daily schooling], pp. 115-125. Göttingen: Hogrefe.

Krug, S. and Lecybyl, R. (2005) 'Die Veränderung von Einstellung, Mitarbeit und Lernleistung im Verlauf einer bezugsnormspezifischen Motivationsintervention' [Changes in attitude, cooperation, and achievement during the course of a reference-norm-specific motivation intervention], in F. Rheinberg (ed.) Motivationsförderung im Schulalltag [Promoting motivation in daily schooling], pp. 95-114. Göttingen: Hogrefe.

Krug, S., Mrazek, J. and Schmidt, C. (1980) 'Motivationsförderung im Sportunterricht durch Leistungsbewertung unter individueller Bezugsnorm' [Promoting motivation in physical education by assessing performance with individual reference norms], Psychologie in Erziehung und Unterricht 27: 278-284.

Léger, L. A., Mercier, D., Gadoury, C. and Lambert, J. (1988) 'The Multistage 20 Metre Shuttle Run Test for Aerobic Fitness', Journal of Sports Sciences 6: 93-101.

Lewis-Beck, M. S., Bryman, A. and Liao, T. F. (2004) The Sage Encyclopedia of Social Science Research Methods. Thousand Oaks, CA: SAGE.

Liu, N. Y.-S., Plowman, S. A. and Looney, M. A. (1992) 'The Reliability and Validity of the 20-Meter Shuttle Test in American Students 12 to 15 Years Old', Research Quarterly for Exercise & Sport 63: 360-365.

Liukkonen, J. and Auweele, Y. V. (2007) 'Prologue', in J. Liukkonen, Y. V. Auweele, B. Vereijken, D. Alfermann and Y. Theodorakis (eds). Psychology for Physical Educators. Student in Focus, 2nd ed., pp. XV-XVIII. Champaign, IL: Human Kinetics.

Lüdtke, O., Köller, O., Marsh, H. W. and Trautwein, U. (2005) 'Teacher Frame of Reference and the Big-Fish-Little-Pond Effect', Contemporary Educational Psychology 30: 263-285.

Margas, N., Fontayne, P. and Brunel, P. C. (2006) 'Influences of Classmates' Ability Level on Physical Self-Evaluations', Psychology of Sport and Exercise 7: 235-247.

Markus, H. and Nurius, P. (1986) 'Possible Selves', American Psychologist 41: 954-969.

Marsh, H. W. (1990) ‘A Multidimensional, Hierarchical Model of Self-Concept: Theoretical and Empirical Justification’, Educational Psychology Review 2: 77-172.

Marsh, H. W. (1993) ‘Physical Fitness Self-Concept: Relations of Physical Fitness to Field and Technical Indicators for Boys and Girls Aged 9-15’, Journal of Sport and Exercise Psychology 15: 184-206.

Marsh, H. W., Chessor, D., Craven, R. G. and 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. and Craven, R. G. (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 Press.

Marsh, H. W. and Craven, R. G. (2002) ‘The Pivotal Role of Frames of Reference in Academic Self-Concept: The “Big-Fish-Little-Pond-Effect”‘, in F. Pajares and T. Urdan (eds) Adolescence and Education, pp. 83-123. Greenwich, CT: Information Age Publishing.

Marsh, H. W. and Craven, R. G. (2006) ‘Reciprocal Effects of Self-Concept and Performance from a Multi-Dimensional Perspective: Beyond Seductive Pleasure and Unidimensional Perspectives’, Perspectives on Psychological Science 1: 133-166.

Marsh, H. W. and Hattie, J. A. (1996) ‘Theoretical Models in Self-Concept’, in B. A. Bracken (ed.) Handbook on Self-Concept, pp. 38-90. Hillsdale, NJ: Erlbaum.

Marsh, H. W. and Hau, K.-T. (2003) ‘Big-Fish-Little-Pond-Effect on Academic Self-Concept: A Cross-Cultural (26-Country) Test of the Negative Effects of Academically Selective Schools’, American Psychologist 58: 364-376.

Marsh, H. W. and Martin, A. J. (2011) ‘Academic Self-Concept and Academic Achievement: Relations and Causal Ordering’, British Journal of Educational Psychology 81: 59-77.

Marsh, H. W. and O'Mara, A. J. (2008) 'Self-Concept Is as Multidisciplinary as It Is Multidimensional. A Review of Theory, Measurement, and Practice in Self-Concept Research', in H. W. Marsh, R. G. Craven and D. M. McInerney (eds) Self-Processes, Learning, and Enabling Human Potential: Dynamic New Approaches, pp. 87-115. Charlotte, NC: Information Age Press.

Marsh, H. W. and 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. and Perry, C. (2005) 'Self-Concept Contributes to Winning Gold Medals: Causal Ordering of Self-Concept and Elite Swimming Performance', Journal of Sport and Exercise Psychology 27: 71-91.

Marsh, H. W., Richards, G. E., Johnson, S., Roche, L. and Tremayne, P. (1994) 'Physical Self-Description Questionnaire: Psychometric Properties and a Multitrait-Multimethod Analysis of Relations to Existing Instruments', Journal of Sport and Exercise Psychology 16: 270-305.

Marsh, H. W., Trautwein, U., Lüdtke, O., Baumert, J. and Köller, O. (2007) 'The Big Fish Little Pond Effect: Persistent Negative Effects of Selective High Schools on Self-Concept after Graduation', American Educational Research Journal 44: 631-669.

Marsh, H. W. and Yeung, A. S. (1998) 'Top-Down, Bottom-up, and Horizontal Models: The Direction of Causality in Multidimensional, Hierarchical Self-Concept Models', Journal of Personality and Social Psychology 75: 509-527.

McCoy, A. R., and Reynolds, A. J. (1998) 'Evaluating implementation', in A. J. Reynolds and H. J. Walberg (eds), Advances in educational productivity (pp. 117-133). Stamford, CT: JAI Press.

McFarlin, D. B., Baumeister, R. F. and Blascovich, J. (1984) 'On Knowing When to Quit: Task Failure, Self-Esteem, Advice, and Non-productive Persistence', Journal of Personality 52: 138-155.

McVeigh, S. K., Payne, A. C. and Scott, S. (1995) 'The Reliability and Validity of the 20-Meter Shuttle Test as a Predictor of Peak Oxygen Uptake in Edinburgh School Children, Age 13 to 14 Years', Pediatric Exercise Science 7: 69-79.

Morgan, K., Sproule, J., Weigand, D. and Carpenter, P. (2005) 'A Computer-Based Observational Assessment of the Teaching Behaviours That Influence Motivational Climate in Physical Education', Physical Education & Sport Pedagogy 10: 83-105.

Mouratidis, A., Lens, W. and Vansteenkiste, M. (2010) 'How You Provide Corrective Feedback Makes a Difference: The Motivating Role of Communicating in an Autonomy-Supporting Way', Journal of Sport & Exercise Psychology 32: 619-637.

O'Mara, A. J., Marsh, H. W., Craven, R. G. and Debus, R. (2006) 'Do Self-Concept Interventions Make a Difference? A Synergistic Blend of Construct Validation and Meta-Analysis', Educational Psychologist 41: 181-206.

Papaioannou, A., Marsh, H. W. and Theodorakis, Y. (2004) 'A Multilevel Approach to Motivational Climate in Physical Education and Sport Settings: An Individual or a Group Level Construct', Journal of Sport and Exercise Psychology 26: 90-118.

Paul, G. and Volk, T. L. (2002) 'Ten Years of Teacher Workshops in an Environmental Problem-Solving Model: Teacher Implementation and Perceptions' The Journal of Environmental Education 33: 10-20.

Paulhus, D. L. (1998) 'Interpersonal and Intrapsychic Adaptiveness of Trait Self-Enhancement: A Mixed Blessing?' Journal of Personality and Social Psychology 74: 1197-1208.

Phillips, D. A. and Zimmerman, M. (1990) 'The Developmental Course of Perceived Competence and Incompetence among Competent Children', in R. J. Sternberg and J.

Kolligian (eds) Competence Considered, pp. 41-66. New Haven, CT: Yale University Press.

Plumert, J. M. (1995) 'Relations between Children's Overestimation of Their Physical Abilities and Accident Proneness', Developmental Psychology 31: 866-876.

Pühse, U. and Gerber, M. (eds). (2005) 'International Comparison of Physical Education'. Aachen: Meyer & Meyer.

Rheinberg, F. (1980) Leistungsbewertung und Lernmotivation [Evaluation of performance and learning motivation]. Göttingen: Hogrefe.

Rheinberg, F. and Krug, S. (2005) Motivationsförderung im Schulalltag [Promoting Motivation in School]. Göttingen: Hogrefe.

Rheinberg, F. and Peter, R. (1982) 'Selbstkonzept, Ängstlichkeit und Schulunlust von Schülern: Eine Längsschnittstudie zum Einfluss des Klassenlehrers' [Students' self-concept, anxiety, and apathy: A longitudinal study on the influence of the class teacher], in F. Rheinberg (ed.) Bezugsnormen zur Schulleistungsbewertung - Analyse und Intervention [Reference norms for evaluating school performance: analysis and intervention], pp. 143-159. Düsseldorf: Schwann.

Robins, R. W. and Beer, J. S. (2001) 'Positive Illusion about the Self: Short-Term Benefits and Long-Term Costs', Journal of Personality and Social Psychology 80: 340-352.

Rosenberg, M. (1965) Society and the Adolescent Self-Image. Princeton, NJ: Princeton University Press.

Rubin, D. B. (1987) Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons.

Schafer, J. L. and Graham, J. W. (2002) 'Missing Data: Our View of the State of the Art', Psychological Methods 7: 147-177.

Schwarzer, R. and Jerusalem, M. (1999) Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der

Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen [Scales for Assessing Teacher and Pupil Characteristics: Documentation of the Psychometric Procedures for the Scientific Evaluation of the Self-Efficacious Schools Model]. Berlin: Freie Universität.

Schwebel, D. C. and Plumert, J. M. (1999) 'Longitudinal and Concurrent Relations among Temperament, Ability Estimation, and Injury Proneness', Child Development 70: 700-712.

Shavelson, R. J. and Bolus, R. (1982) 'Self-Concept: The Interplay of Theory and Methods', Journal of Educational Psychology 74: 3-17.

Shavelson, R. J., Hubner, J. J. and Stanton, G. C. (1976) 'Self-Concept: Validation of Construct Interpretations', Review of Educational Research 46: 407-441.

Solberg Nes, L., Segerstrom, S. C. and Sephton, S. E. (2005) 'Engagement and Arousal: Optimism's Effects During a Brief Stressor', Personality and Social Psychology Bulletin 31: 111-120.

Spence, J. C., McGannon, K. R. and Poon, P. (2005) 'The Effect of Exercise on Global Self-Esteem: A Quantitative Review', Journal of Sport and Exercise Psychology 27: 311-334.

Stiller, J., Würth, S. and Alfermann, D. (2004) 'Die Messung des physischen Selbstkonzepts (PSK). Zur Entwicklung der PSK-Skalen für Kinder, Jugendliche und junge Erwachsene [Measuring the Physical Self-Concept (PSK). On the Development of PSK-Scales for Children, Adolescents and Young Adults]', Zeitschrift für Differentielle und Diagnostische Psychologie 25: 239-257.

Suls, J., Martin, R. and Wheeler, L. (2002) 'Social Comparison: Why, with Whom, and with What Effect?' Current Directions in Psychological Science 11: 159-163.

Taylor, S. E. and Brown, J. D. (1988) 'Illusion and Well-Being: A Social Psychological Perspective on Mental Health', Psychological Bulletin 103: 193-210.

Taylor, S. E., Lerner, J. S., Sherman, D. K., Sage, R. M. and McDowell, N. K. (2003) ‘Portrait of the Self-Enhancer: Well Adjusted and Well Liked or Maladjusted and Friendless?’ Journal of Personality and Social Psychology 84: 165-176.

Tracey, D. K., Marsh, H. W. and Craven, R. G. (2003) ‘Self-Concepts of Preadolescents with Mild Intellectual Abilities. Issues of Measurement and Educational Placement’, in H. W. Marsh, R. G. Craven and D. M. McInerney (eds) International Advances in Self-Research, pp. 203-229. Greenwich: Information Age Publishing.

Vancouver, J. B., Thompson, C. M., Tischner, E. C. and Putka, D. J. (2002) ‘Two Studies Examining the Negative Effect of Self-Efficacy on Performance’, Journal of Applied Psychology 87: 506-516.

Vealey, R. S. (2002) ‘Personality and Sport Behaviour’, in T. S. Horn (ed) Advances in Sport Psychology, pp. 43-82. Champaign, IL: Human Kinetics.

Valentine, J. C., DuBois, D. L. and Cooper, H. (2004) ‘The Relation between Self-Beliefs and Academic Achievement: A Meta-Analytic Review’, Educational Psychologist 39: 111-133.

von Collani and Herzberg, P. Y. (2003) ‘Eine revidierte Fassung der deutschsprachigen Skala zum Selbstwertgefühl von Rosenberg‘ [A Revised Version of the German Adaptation of Rosenberg's Self-Esteem Scale], Zeitschrift für Differentielle und Diagnostische Psychologie 24: 3-7.

Walsh, J. P., Henderson, C. M. and Deighton, J. (1988) ‘Negotiated Belief Structures and Decision Performance: An Empirical Investigation’, Organizational Behavior and Human Decision Processes 42: 194-216.

Weinberg, R. S. and Gould, D. (2007) Foundations of Sport and Exercise Psychology, 4th edn. Champaign, IL: Human Kinetics.

Weiner, B. (2005) 'Motivation from an Attribution Perspective and the Social Psychology of Perceived Competence', in A. J. Elliot and C. S. Dweck (eds) Handbook of Competence and Motivation, pp. 73-84. New York: Guilford Press.

Weiss, M. R. and Amorose, A. J. (2008) 'Motivational Orientations and Sport Behaviour', in T. S. Horn (ed) Advances in Sport Psychology, pp. 115-155. Champaign, IL: Human Kinetics.

Weiss, M. R. and Horn, T. S. (1990) 'The Relationship between Children's Accuracy Estimates of Their Physical Competence and Achievement-Related Behaviors', Research Quarterly for Exercise and Sport 61: 250-258.

Welk, G. J. and Eklund, B. (2005) 'Validation of the Children and Youth Physical Self Perceptions Profile for Young Children', Psychology of Sport & Exercise 6: 51-65.

Wrosch, C., Scheier, M. F., Miller, G. E., Schulz, R. and Carver, C. S. (2003) 'Adaptive Self-Regulation of Unattainable Goals: Goal Disengagement, Goal Reengagement, and Subjective Well-Being', Personality and Social Psychology Bulletin 29: 1494-1508.

Wyss, T., Marti, B., Rossi, S., Kohler, U. and Mäder, U. (2007) 'Assembling and Verification of a Fitness Test Battery for the Recruitment of the Swiss Army and Nation-Wide Use', Schweizerische Zeitschrift für Sportmedizin und Sporttraumatologie 55: 126-131.

Yamnill, S., & McLean, G. N. (2001) 'Theories supporting transfer of training' Human Resource Development Quarterly 12: 195-208

Zeidner, M. and Schleyer, E. J. (1999) 'The Big-Fish-Little-Pond Effect for Academic Self-Concept, Test Anxiety, and School Grades in Gifted Children', Contemporary Educational Psychology 24: 305-329.

Table 1
Descriptive statistics for the study variables and results of preliminary analyses of both intervention and control groups

Variable	Intervention group			Control group			Equivalence of conditions			Sex differences		
	Boys (n = 126)	Girls (n = 120)	Total (n = 246)	Boys (n = 120)	Girls (n = 98)	Total (n = 218)	t	p	d	t	p	d
Pre												
SC of endurance	2.89 (.62)	2.55 (.50)	2.72 (.55)	2.99 (.66)	2.52 (.57)	2.77 (.60)	1.29	.20	.09	7.18	<.0005	.60
Actual endurance	387 (126)	325 (114)	357 (123)	395 (115)	316 (122)	359 (126)	.22	.83	.02	6.21	<.0005	.54
SC of strength	3.09 (.54)	2.80 (.53)	2.95 (.56)	3.10 (.47)	2.78 (.57)	2.95 (.55)	.10	.92	.00	5.96	<.0005	.52
Actual strength	3.31 (.49)	3.08 (.43)	3.20 (.47)	3.36 (.46)	3.06 (.38)	3.23 (.46)	.59	.55	.06	6.37	<.0005	.64
General SC	3.44 (.51)	3.30 (.49)	3.37 (.50)	3.48 (.39)	3.35 (.49)	3.42 (.44)	1.24	.22	.10	2.99	.003	.26
Individualized TFR	3.03 (.67)	3.04 (.52)	3.04 (.60)	3.03 (.57)	3.07 (.58)	3.05 (.57)	.24	.81	.02	.40	.69	.04
Post												
SC of endurance	3.02 (.64)	2.65 (.55)	2.84 (.62)	2.98 (.66)	2.53 (.64)	2.77 (.69)						
Actual endurance	414 (112)	362 (122)	389 (119)	427 (107)	358 (128)	396 (123)						
SC of strength	3.14 (.58)	2.85 (.57)	3.00 (.59)	3.12 (.45)	2.82 (.55)	2.98 (.53)						
Actual strength	3.39 (.48)	3.08 (.49)	3.24 (.50)	3.37 (.49)	3.10 (.41)	3.25 (.48)						
General SC	3.41 (.59)	3.33 (.56)	3.37 (.58)	3.46 (.44)	3.37 (.52)	3.42 (.48)						
Individualized TFR	3.11 (.66)	3.10 (.53)	3.11 (.60)	3.03 (.57)	3.13 (.63)	3.07 (.60)						

Note. Group means with standard deviations in parentheses. SC = self-concept. Actual endurance = test score in the 20-meter shuttle run test (in seconds). Actual strength = test score in the shot-put performance test (in meters). Individualized TFR = Individualized teacher frame of reference.

Table 2
Descriptive statistics for the variable veridicality of the self-concept of endurance and the self-concept of strength, with results of the ANOVA with repeated measures for the three subgroups

	Veridicality of the Self-Concept of Endurance								
	Underestimators			Realists			Overestimators		
	Int (n=38)	Cont (n=26)		Int (n=176)	Cont (n=165)		Int (n=32)	Cont (n=27)	
Pre	-1.53 (.49)	-1.40 (.31)		.06 (.49)	-.01 (.51)		1.47 (.38)	1.43 (.41)	
Post	-.67 (.84)	-1.33 (.56)		.03 (.75)	-.01 (.72)		.69 (.81)	1.22 (.58)	
	ANOVA with repeated measures								
Source	F-ratio	η^2	p	F-ratio	η^2	p	F-ratio	η^2	p
Time	21.26*	.262	<.0005	.05	.000	.831	24.81*	.303	<.0005
Group	3.13	.051	.082	.69	.002	.406	2.89	.048	.094
Time x Group	15.05*	.201	<.0005	.59	.002	.441	8.06*	.124	.003
	Veridicality of the Self-Concept of Strength								
	Underestimators			Realists			Overestimators		
	Int (n=45)	Cont (n=42)		Int (n=157)	Cont (n=135)		Int (n=44)	Cont (n=41)	
Pre	-1.56 (.07)	-1.63 (.08)		.04 (.05)	-.04 (.05)		1.60 (.07)	1.69 (.07)	
Post	-.99 (.13)	-1.39 (.14)		.03 (.06)	-.08 (.07)		1.01 (.10)	1.53 (.10)	
	ANOVA with repeated measures								
Source	F-ratio	η^2	p	F-ratio	η^2	p	F-ratio	η^2	p
Time	24.99*	.221	<.0005	.33	.001	.563	32.08*	.276	<.0005
Group	3.36	.037	.070	1.54	.005	.215	9.07*	.094	.003
Time x Group	4.16*	.045	.022	.17	.001	.684	10.30*	.109	.001

Note. Group means are presented with standard deviations in parentheses. Int = intervention group, Cont = control group.

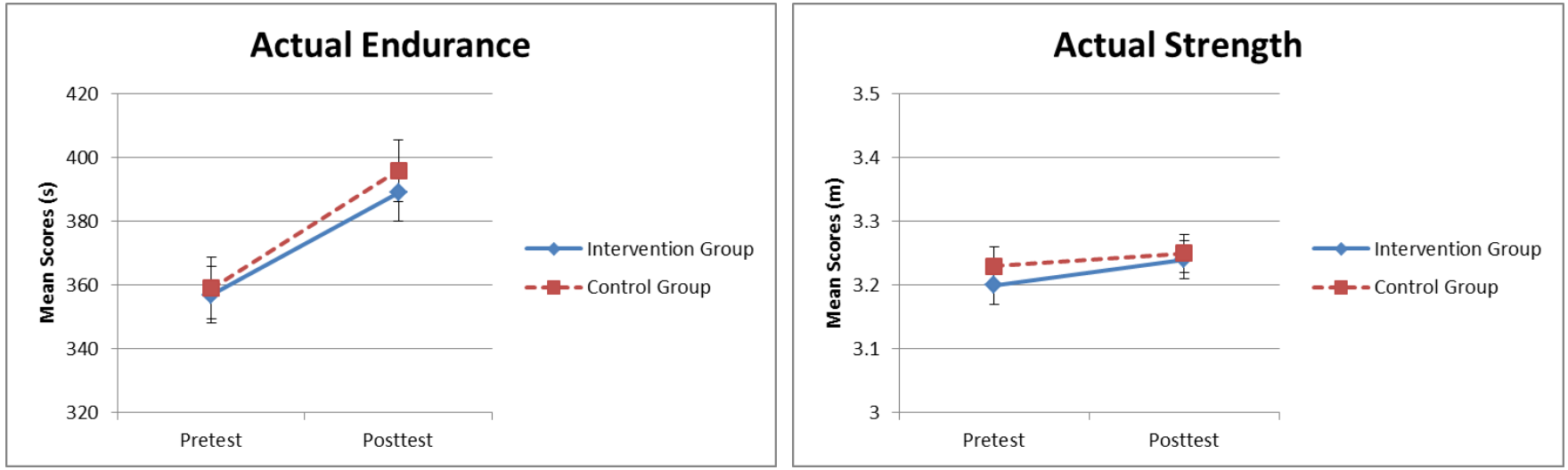

Figure 1 Means of actual endurance and means of actual strength for the intervention and control group. Error bars represent standard error of the mean.

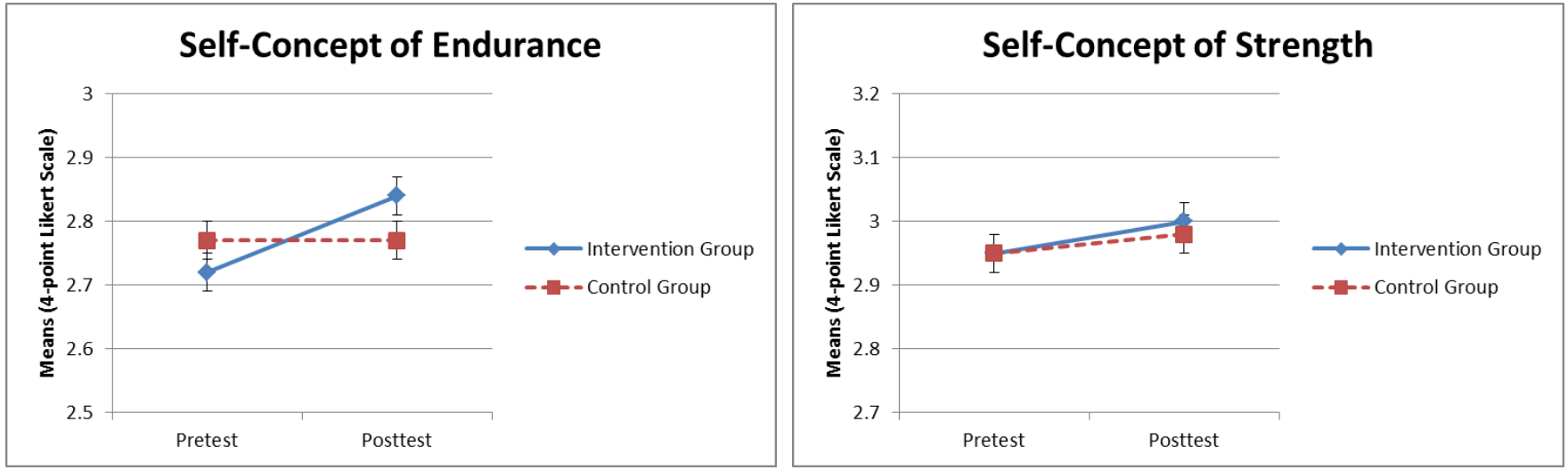

Figure 2 Means of the self-concept of endurance and means of the self-concept of strength for the intervention and control group. Error bars represent standard error of the mean.

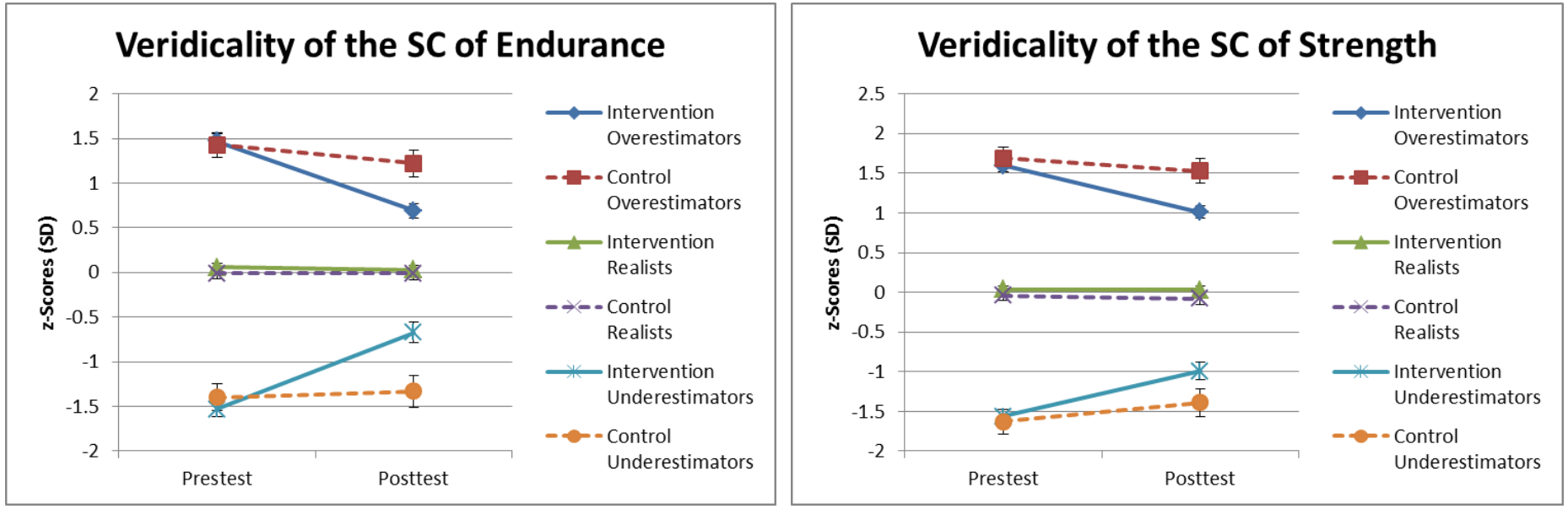

Figure 3 Means of veridicality (= difference scores between z scores of self-concept and z scores of actual physical competence) of the self-concept of endurance and the self-concept of strength for the intervention and control group. Error bars represent standard error.