A multidimensional physical self-concept: A construct validity approach to theory, measurement and research

HERBERT W. MARSH
University of Western Sydney, Australia

ABSTRACT

Interest in physical self-concept stems from its recognition as a valued outcome, its role as a moderator variable, interest in its relation with other constructs, and concerns with methodological and measurement issues. My purpose is to provide an overview of my physical self-concept research and the construct validation approach that has guided it: the construction of a multidimensional physical self-concept instrument based on theory and research, its psychometric evaluation (reliability and confirmatory factor analysis), tests of convergent and divergent validity, validation in relation to external criteria, and application to substantive research issues and practice. From a construct validation perspective, theory, measurement, empirical research, and practice are inexorably intertwined so that the neglect of one will undermine the others. The strongest contribution of my physical self-concept research may be the development of instruments, based on strong empirical and theoretical foundations, for the measurement of multiple dimensions of physical self-concept. The research also demonstrates a research program based on a construct validity approach in which an emphasis on good measurement is a critical feature of good research. This approach should be useful to other areas of sport psychology and to sports sciences more generally. No longer can sport/exercise psychologists simply pull together an *ad hoc* set of items that are more or less related to the construct of interest and claim - with any credibility - that they have a new instrument.

Key words: Construct validity, Physical fitness, Physical self-concept.

Introduction

The focus of my paper is on a construct validation approach to the measurement of a multidimensional physical self-concept. For some time there seems to have been general agreement among sport/exercise psychology researchers for the need to develop sport specific instruments and to evaluate them within a construct validity framework (Nelson, 1989). Vealey (1986) claimed that significant advances in sport/exercise psychology research "await sport-specific conceptualization and measurement instrumentation" (p. 222). In his review of sport and exercise tests Ostrow (1990) reported substantial gains over the last 25 years, but emphasized that many tests are still "one shot assessments," lacking further development and refinement. More specifically, Gill, Dzewaltowski, and Deeter (1988) concluded that "Within sport

Note: I would like to acknowledge the financial support from the Australian Research Council. This paper was the basis of an invited keynote address at 10th *World Congress on Sport Psychology*, Skiathos, Greece, 28 May - 2 June 2001.

Address: Herbert W. Marsh, Director, SELF Research Centre, University of Western Sydney, Bankstown Campus, Penrith, New South Wales 1797, Australia. E-mail: h.marsh@uws.edu.au

psychology the most promising work on individual differences involves the development and use of sport-specific constructs and measures" (pp. 139-140) and argued for the construction of multidimensional instruments based on theory, followed by item and reliability analysis, exploratory and confirmatory factor analysis, tests of convergent and divergent validity, validation in relation to external criteria, and application in research and practice.

Let me summarise my own idiosyncratic perspectives of this development. Despite recognition of the importance of developing reliable and valid measures, it is evident that the quality of measures in sport/exercise research has been weak. In a monumental effort to catalogue sport/exercise measures used over a 25-year period, Andrew Ostrow (1990) developed the Directory of Psychological Tests in the Sport and Exercise Sciences. He included all instruments from the published sport/exercise literature with reliability or validity information. In addition to the valuable information about individual measures, the Directory also provided a barometer for evaluating the quality of measurement in our field. Indeed, one of the expressed intents of the Directory was to force researchers, test authors, reviewers, journal editors, publishers, and test consumers to embrace a higher standard of measurement quality. However, of the 175 instruments summarised in the Directory, only 1/3 had items based on a conceptual or theoretical framework, less than 1/4 reported factor analyses, and less than 10% showed evidence of extensive reference support. Apparently, measurement practice in the 25 years prior to 1990 did not measure up to the high ideals espoused by leading researchers. However, there has been much progress in measurement sophistication in sport/exercise psychology in the last decade or so. Some of this progress is highlighted, for example, in Professor Joan Duda's Advance in Sport/Exercise Psychology Measurement and in Professor Andrew Ostrow's new Directory of Psychological Tests in Sport and Exercise Science. This progress is the result of: (a) more carefully developed instruments, (b) better articulation of the links between instrument design, theory, and practice, and (c) improved application of methodological and statistical techniques. This improvement reflects the higher standards expected of sport/exercise researchers. Thankfully, the hey day of the "one shot" instruments seems to have ended. No longer can sport/exercise psychologists simply pull together an ad hoc set of items that are more or less related to the construct of interest and claim – with any credibility – that they have a new instrument.

A construct validation approach

From a construct validation perspective, theory, measurement, empirical research, and practice are inexorably intertwined so that the neglect of one will undermine the others (see Marsh, 1997, for further discussion). Validation from this perspective seeks to assess the usefulness of interpretations based on responses to the measure, not to establish their absolute truth or reality. Ideally, the validation is an on-going process in which theory and practice are used to develop a measure, empirical research is used to test the theory and the measure, both the theory and the measure are revised in relation to research, new research is conducted to test these refinements, and theory and research are used to inform practice. Reality seldom matches this ideal. All too often in the not so distant past, measures in the sports sciences (Ostrow, 1990) and other social science disciplines were largely *ad hoc* or "one shot" endeavors that were not soundly based on theory, not systematically evaluated, and not refined on the basis of subsequent theoretical or substantive developments. Weak measures undermine research and theory evaluation, thereby limiting their contribution to practice. Thus, even though Ostrow (1990) reported substantial gains in sport/

exercise measurement over the last 25 years, he was forced to concede that many tests were still "one shot assessments," lacking further development and refinement.

It can be argued that all constructs in sport/exercise psychology are hypothetical constructs and so must be validated using a construct validity approach. However, many of our constructs suffer in that "everybody knows what it is" so that many researchers do not feel compelled to provide appropriate theoretical definition of what they are measuring or to evaluate fully the psychometric properties of responses to their measures. Because our constructs are hypothetical constructs, their usefulness must be established by investigations of their construct validity. Although this approach is most typically applied to psychological tests, the logic of construct validation also applies to other measurement techniques such as interviews, surveys, behavioral observations, and physiological measures. For example, even a construct as apparently tangible as "body fat" is a hypothetical construct, as can be readily seen in the diverse and only partly consistent ways that this construct is inferred. Although researchers may claim that their approach to inferring body fat is the "gold standard", support for such claims must be based on a construct validation approach rather than the sophistication and cost of their technical equipment. "High tech" measures have to be good in relation to typical standards of construct validity and in comparison to more easily obtained measures before they earn the label of "gold standard". From a construct validation approach, theory, measurement, empirical research, and practice are inexorably intertwined so that the neglect of one will undermine the others.

In this paper, I will demonstrate a construct validity approach in my self-concept research program. Construct validity investigations can be classified as within-network or between-network studies.

- Within-network studies explore the internal structure of a construct. They typically employ empirical techniques such as factor analysis or multitrait-multimethod (MTMM) analysis.
- Between-network studies attempt to establish a logical, theoretically consistent pattern of relations between measures of a construct and other constructs.

The construct validity approach ideally incorporates logical, correlational, and experimental approaches to evaluate the validity of a construct.

1. Logical analysis examines the logical consistency of the construct definition, measurement instruments (instructions, item format, scoring procedures, etc.), and predictions.

2. Correlational techniques can be used to investigate the (within-network) structure of self-concept and the (between-network) relations between self-concept and other constructs. Typically, construct validation research involves showing that multiple indicators of the same construct are more closely related to each other than to indicators of different constructs. Useful statistical tools include factor analysis, multitrait-multimethod analysis, and path analysis.

3. Experimental techniques are also useful in testing the validity of interpretations of self-concept responses. Thus, for example, theory may suggest that a certain intervention should enhance academic self-concept. To the extent that the intervention leads to the enhancement of self-concept, then there is support for the theory and the procedure used to measure self-concept. A potentially useful test of the multidimensionality of self-concept is to test whether the intervention influences those facets of self-concept most relevant to the intervention, but also to ascertain that the intervention does not influence, or has substantially less influence on, those facets of self-concept that it is not intended to influence (e.g., Marsh, Richards, & Barnes, 1986a, 1986b).

Self-concept

A positive self-concept is valued as a desirable outcome in many disciplines such as sport, health, educational, developmental, clinical, and social psychology. Self-concept is frequently posited as a mediating variable that facilitates the attainment of other desired outcomes such as physical activity, exercise adherence or health-related physical fitness. Researchers with a major focus on other constructs are often interested in how constructs in their research are related to self-concept. Methodologists are also concerned with particular measurement and methodological issues inherent in the study of self-concept. Even in studies where self-concept is not the major focus of interest, it is useful to evaluate self-concept because of its importance as a mediating variable that facilitates the attainment of other desired outcomes (e.g., physical activity, exercise adherence, health-related physical fitness). Thus, for example, if an intervention unintentionally undermines self-concept then it is unlikely to have long lasting effects on its intended outcomes. In contrast, if an intervention enhances self-concept as well as its desired outcome (e.g., physical skills or fitness), then the effects are more likely to be long lasting.

Interest in self-concept has a long and controversial history and it is one of the oldest areas of research in the social sciences. The longest chapter in William James' 1890 textbook, the first introductory textbook in psychology, was devoted to self-concept and introduced many issues of current relevance. Despite the rich beginning provided by William James, advances in theory, research and measurement of self-concept were slow during the hey day of behaviorism. It is only in the last 25 years that there has been a resurgence in self-concept research. Particularly prior to the 1980s, reviewers (e.g., Burns, 1979; Shavelson, Hubner, & Stanton, 1976; Wells & Marwell, 1976; Wylie, 1974, 1979) typically emphasized the lack of theoretical basis in most studies, the poor quality of measurement instruments used to assess self-concept, methodological shortcomings, and a general lack of consistent findings. Similar observations led Hattie (1992) to describe this period as one of "dustbowl empiricism" in which the predominant research design in self-concept studies was "throw it in and see what happens."

Self-concept, like many other psychological constructs, suffers in that "everybody knows what it is", so that many researchers do not feel compelled to provide any theoretical definition of what they are measuring. Because self-concept is a hypothetical construct, its usefulness must be established by within-network and between-network studies of its construct validity. Within-network studies explore the internal structure of self-concept. They test, for example, the dimensionality of self-concept and may seek to show that the construct has consistent, distinct multidimensional components (e.g., physical, social, academic self-concept). These studies typically employ empirical techniques such as factor analysis or multitrait-multimethod (MTMM) analysis (Marsh, 1988). Between-network studies attempt to establish a logical, theoretically consistent pattern of relations between measures of self-concept and other constructs. The resolution of at least some within-construct issues should be a logical pre-requisite to conducting between-construct research.

Research prior to the 1980s had made limited progress toward resolving either the within- or between-construct issues. In fact, most research was directed towards the between-construct issues of relating self-concept to other constructs, whereas insufficient attention had been given to the within-construct issues that should have been the basis of constructing appropriate measurement instruments. In retrospect, this emphasis on between-construct research to the exclusion of within-construct research may have been counter-productive and appears to be one reason why findings were not more consistent across different studies. In contrast, studies since

the 1980s have made important advances in theory, measurement and research that have important implications for practice. One cornerstone for these advances is the classic review article by Shavelson et al. (1976). This review is central to this chapter in that it provided a theoretical blueprint for the development of the many self-concept instruments considered here (see Marsh & Hattie, 1996 for a more detailed evaluation of the Shavelson model and its impact on subsequent self-concept research).

The Shavelson model

Shavelson et al. (1976) noted critical deficiencies in self-concept research including inadequate definitions of the self-concept construct, a dearth of appropriate measurement instruments, and the lack of rigorous tests of counter-interpretations. They concluded that "it appears that self-concept research has addressed itself to substantive problems before problems of definition, measurement, and interpretation have been resolved" (p. 470). However, unlike many other reviews, Shavelson et al. emphasized "our approach is constructive in that we: (a) develop a definition of self-concept from existing definitions, (b) review some steps in validating a construct interpretation of a test score, and (c) apply these steps in examining five popularly used self-concept instruments" (p. 470).

Shavelson et al. (1976) began their review by developing a theoretical definition of self-concept. An ideal definition, they emphasized, should consist of the nomological network containing within-network and between-network components. The within-network portion of the network pertains to specific features of the construct – its components, structure, and attributes and theoretical statements relating these features. The between-network portion of the definition locates the construct in a broader conceptual space, indicating how self-concept is related to other constructs. Thus, for example, dividing self-concept into academic, social and physical components is a within-network proposition whereas a related between-network proposition is that physical self-concept is more strongly related to physical fitness than are academic and social self-concepts.

Construct definition of self-concept. Shavelson et al. (1976) integrated features of various definitions of self-concept to form their working definition of self-concept that was then used to integrate empirical evidence. According to their definition, self-concept is a person's self-perceptions that are formed through experience with and interpretations of one's environment. They are influenced especially by evaluations by significant others, reinforcements, and attributions for one's own behavior. According to Shavelson et al., self-concept is not an entity within the person, but a hypothetical construct that is potentially useful in explaining and predicting how a person acts. These self-perceptions influence the way one acts and these acts in turn influence one's self-perceptions. Consistent with this perspective, Shavelson et al. noted that self-concept is important as both an outcome and as a mediating variable that helps to explain other outcomes. Shavelson et al. also distinguished between self-concepts based on a person's own self-perceptions and inferred self-concepts that are based on inferences by another person, noting that they would focus on the former. Shavelson et al. identified seven features that were critical to their definition of the self-concept construct:

- It is organized or structured, in that people categorize the vast amount of information they have about themselves and relate these categories to one another.
- It is multifaceted, and the particular facets reflect a self-referent category system adopted by a particular individual and/or shared by a group.
- It is hierarchical, with perceptions of personal behavior in specific situations at the

base of the hierarchy, inferences about self in broader domains (e.g., social, physical, and academic) at the middle of the hierarchy, and a global, general self-concept at the apex.

- The hierarchical general self-concept – the apex of the hierarchy – is stable, but as one descends the hierarchy, self-concept becomes increasingly situation specific and, as a consequence, less stable. There are reciprocal relations between self-concept at each level in that self-perceptions at the base of the hierarchy may be attenuated by conceptualizations at higher levels, and changes in general self-concept may require changes in many situation-specific instances.
- Self-concept becomes increasingly multifaceted as the individual moves from infancy to adulthood.
- Self-concept has both a descriptive and an evaluative aspect such that individuals may describe themselves ("I am happy") and evaluate themselves ("I do well in mathematics"). Evaluations can be made against some absolute ideal (the five-minute mile), a personal, internal standard (a personal best), a relative standard based on comparisons with peers, or the expectations of significant others. Individuals may differentially weight specific dimensions.
- Self-concept can be differentiated from other constructs. Thus, for example, academic and physical self-concepts can be differentiated from other constructs such as academic achievement and physical fitness respectively.

Shavelson et al. (1976) also presented one possible representation of this hierarchical model in which General-self appeared at the apex and was divided into academic and nonacademic self-concepts at the next level. Academic self-concept was further divided into self-concepts in particular subject areas (e.g., mathematics, English, etc.). Nonacademic self-concept was divided into three areas: Social self-concept which was subdivided into relations with peers and with significant others; Emotional self-concept; and Physical self-concept which was subdivided into physical ability and physical appearance. This model posits a structure of self-concept that resembles British psychologists' hierarchical model of intellectual abilities where general ability (like Spearman's "g") was at the apex.

In his review, Shavelson et al. (1976) emphasized a construct validity approach to the self-concept research and illustrated this approach in an evaluation of the construct validity of five popular self-concept instruments: Brookover's Self-concept of Ability Scale; Coopersmith's Self-Esteem Inventory; Gordon's How I See Myself Scale; the Piers-Harris Children's Self-concept Scale; Sear's Self-concept Inventory. Whereas the Brookover and Coopersmith instruments do not purport to measure a physical component of self-concept, Shavelson et al. noted some evidence for a physical component in each of the other instruments. Shavelson et al. reported that a factor analysis of the Gordon instrument identified a physical appearance factor, although Gordon did not specifically design the instrument to measure separate components of self. The Piers-Harris was originally intended to measure overall self-concept, but factor analyses have identified different clusters of items including one related to physical appearance. Shavelson et al., however, concluded that there was insufficient evidence to support the interpretability of separate scale scores (but see MTMM study by Marsh, 1990a). Sears included a physical ability scale in her original 100-item instrument, but the most analogous scale on the revised instrument was her attractive appearance scale. Shavelson et al. reviewed evidence on discriminant validity for this instrument, but found little support for a physical factor. In summary, none of the five instruments reviewed by Shavelson et al. provides a clearly interpretable measure of physical self-concept that can be differentiated from general self-concept and other domains of self-concept. From a practical perspective, these older instruments apparently are not very useful for sport/exercise psychologists interested in measuring physical self-concept.

It is important to emphasize that at the time

Shavelson et al. first developed their model there was only modest support for the hypothesized domains and no one instrument considered in their review was able to differentiate among even the broad academic, social, and physical domains. In this respect, the Shavelson et al. model provided a theoretical model for the development of new theory, measurement, and research. In order to address these concerns, the Self-Description Questionnaire (SDQ) instruments were developed for preadolescent primary school students (SDQI), adolescent high school students (SDQII), and late adolescents and young adults (SDQIII). Reviews of subsequent SDQ research (Boyle, 1994; Byrne, 1984, 1996b; Hattie, 1992; Marsh, 1990b, 1990c, 1993a; Marsh & Shavelson, 1985; Wylie, 1989) supported the multifaceted structure of self-concept and demonstrated that self-concept cannot be adequately understood if its multidimensionality is ignored. The three SDQ instruments have provided particularly strong tests of the Shavelson et al. model, and have been evaluated to be among the best multidimensional instruments in terms of psychometric properties and construct validation research (Boyle, 1994; Byrne, 1984, 1996b; Hattie, 1992; Wylie, 1989). Here I review the development of the SDQ instruments with a particular emphasis on physical self-concept and its relevance to sport and exercise psychology.

Self Description Questionnaires and physical self-concept

Historical background

It is relevant to set a historical stage for discussing early development of the SDQ instruments. At the time I began my work in this area, researchers seriously argued that self-concept was either unidimensional or that the so called multiple dimensions of self-concept were so highly correlated that they could not be adequately distinguished (see review by Marsh & Hattie, 1996). The self-concept facets proposed in the Shavelson model, as well as their hypothesized structure, were heuristic and plausible, but they were not empirically validated by existing research. At the time, Shavelson et al. were unable to identify any instrument for measuring multiple facets of self-concept as posited in their model. Existing self-concept instruments typically consisted of a hodgepodge of self-referent items. Whereas factor analyses of responses to such instruments usually resulted in more than one factor, the factors were typically not replicable, easily interpreted, or consistent with the design of the instrument (see Marsh & Smith, 1982; Shavelson et al., 1976; Wylie, 1974, 1979). Empirical research based on these instruments led some researchers (e.g., Coopersmith, 1967; Marx & Winne, 1978) to argue that the facets of self-concept were so heavily dominated by a general factor that they could not be adequately differentiated. Based on this early self-concept research following the Shavelson model, Byrne (1984, pp. 449-450) noted that "Many consider this inability to attain discriminant validity among the dimensions of SC [self-concept] to be one of the major complexities facing SC researchers today."

Hence, I began my research in a historical context in which the existence of a multidimensional self-concept was questioned. From this perspective, I reasoned that the determination of whether theoretically consistent and distinguishable facets of self-concept exist, and their content and structure if they do exist, should be prerequisite to the study of how these facets, or overall self-concept, are related to other variables. In adopting such an approach, atheoretical and/or purely empirical approaches to developing and refining measurement instruments were rejected. Instead, an explicit theoretical model was taken to be the starting point for instrument construction, and empirical results were used to support, refute or revise the instrument *and* the theory upon which it is based. In applying this ap-

proach, the Shavelson et al. model was judged the best available theoretical model of self-concept. Implicit in this approach is the presumption that theory building and instrument construction are inexorably intertwined, and that each will suffer if the two are separated. In this sense, the SDQ instruments are based on a strong empirical foundation and a good theoretical model. Consistent with this approach, SDQ research described below provided support for the Shavelson et al. model, but also led to its subsequent revision.

Within-network concerns: The structure of self-concept

In this early SDQ research, I critically evaluated the within-network components of the Shavelson et al. model and the psychometric properties of the SDQ instruments. SDQ scales were posited on the basis of the Shavelson et al. model, item pools were constructed for each scale, and factor analyses and item analyses were used to select and refine the items eventually used to represent each scale. The internal consistency of the scales from the three SDQ instruments was good – typically in the .80s and .90s. The stability of SDQ responses was also good, particularly for older children. For example, the stability of SDQIII scales measured on four occasions varied from a median of .87 for a one-month interval to a median of r = .74 for intervals of 18 months or longer. Dozens of factor analyses by diverse samples differing in gender, age, country, and language have consistently identified the factors that each SDQ instrument is designed to measure. Marsh (1989; also see Marsh, 1990b) summarized factor analyses of more than 12,000 sets of responses from the normative archives of the three SDQ instruments. In addition to clearly identifying all of the factors that each of the three SDQ instruments are designed to measure, the results indicate that the domains of self-concept are remarkably distinct (median *r*s among the SDQ scales vary between .1 and .2 for the three SDQ instruments). Hence, the correlations among the different SDQ factors were so low as to call into question the usefulness of a hierarchical or global self-concept. These extremely small correlations are also in marked contrast to earlier conclusions that self-concept was unidimensional or that the factors were so highly correlated that they could not be distinguished. In retrospect, it seems that empirical results based on poor measures and, perhaps, misinterpretation of statistical analyses had led researchers to inappropriate conclusions (see Marsh & Hattie, 1996). This juxtaposition between results based on the SDQ instruments and the historical context from which this research grew provides a dramatic testimonial for the relevance of the construct validation approach underpinning SDQ research.

Whereas SDQ results provide strong support for the Shavelson et al. model and the multidimensionality of self-concept, they also posed some complications. The strong hierarchical structure posited by Shavelson et al. required self-concepts to be substantially correlated, but the small sizes of correlations actually observed implied that any hierarchical structure of the self-concept responses must be much weaker than anticipated. More specifically, in the Shavelson et al. model Math and Verbal self-concepts were assumed to be correlated substantially so that they can be described in terms of a single higher order academic self-concept. Factor analyses, however, resulted in correlations between Verbal and Math self-concepts that were close to zero. Complications such as these led to the Marsh/Shavelson revision (Marsh & Shavelson, 1985; Marsh, Byrne & Shavelson, 1988) of the original Shavelson et al. model. It is important to emphasize, however, that this revision of the model should be viewed as a strength of the research program. In an active research program – particularly in its early stages – theory should be dynamic so that it grows with concurrent developments in measurement, research, and practice.

The Shavelson et al. and Marsh/Shavelson models prompted between-network research primarily in the academic domains of self-concept (see Marsh, 1990b, 1990c, 1993a; Marsh & Craven, 1997; for overviews of research summarized here). Thus, for example, academic achievement is nearly uncorrelated with general and nonacademic domains of self-concept but is substantially related to academic self-concept. As students grow older, the various domains of self-concept become more differentiated (less correlated) and self-concept becomes more predictable from external criteria and from the evaluations of significant others (teachers, parents, and peers). Interventions designed to enhance academic achievement and academic self-concept have substantially stronger effects on academic self-concept than on nonacademic and general areas of self-concept (e.g., Marsh, 1990b, 1993a; Marsh & Craven, 1997). In longitudinal panel studies, academic self-concept contributed to subsequent school grades beyond the contribution of standardized test scores (e.g., IQ) and prior school grades, suggesting that academic self-concept is causally related to academic achievement as well as being correlated with academic achievement. Academic self-concept and self-concepts in particular school subjects are also related to other academic outcomes such as time spent on homework, academic course selection, and subsequent university attendance. Marsh and Yeung (1997) demonstrated that when academic achievement and academic self-concept in specific school subjects were used to predict subsequent coursework, academic achievement did not contribute beyond what could be explained by academic self-concept. Because self-concept has such a strong effect on choices about what individuals will pursue, it is not surprising that self-concept is also related to subsequent accomplishments that follow from these choices.

In order to more fully test the academic component of the Marsh/Shavelson model, Marsh (1990c; 1993a; Marsh & Craven, 1997) subsequently developed the Academic SDQ instrument that measured a larger number of more specific domains of academic self-concept. Factor analyses demonstrated that students were able to differentiate academic self-concepts related to 15 school subjects and that relations with parallel measures of academic achievement further supported this level of differentiation. Although hierarchical confirmatory factor analyses (HCFAs) demonstrated that relations among core academic subjects could be explained by two higher-order factors (Math/academic and Verbal/academic), much reliable variance in the more specific domains could not be explained in terms of the higher order constructs. Based on this research, Marsh (1990b, 1990c, 1993a) argued that self-concept researchers and practitioners should measure self-concept at a level of specificity consistent with their particular issues as well as, perhaps, more general measures of academic self-concept and esteem that are typical in most self-concept research. Marsh (1993a) also noted that the development of a new instrument specific to one domain was consistent with the Shavelson et al. model, and this logic was also the basis for the development of the Physical SDQ (PSDQ) instrument described later in this chapter.

Relations between SDQ Physical Self-concepts and other self-concept domains

SDQ research provides good support for the construct validity of the Physical Appearance and Physical Ability scales that appear on the SDQI, SDQII, and SDQIII instruments. In particular, consistent with the emphasis on the multidimensional perspective to self-concept research, SDQ research shows that Physical Ability and Physical Appearance Self-concepts are distinct components of self-concept (see Table 1). Whereas Physical Ability and Physical Appearance Self-concepts are modestly related to

Table 1
Correlations between physical self-concept scales and other areas of self-concept for each of the SDQ instruments

SDQ scales	SQD instruments					
	SDQI Physical		SDQII Physical		SDQIII Physical	
	Ability	Appr	Ability	Appr	Ability	Appr
Physical Ability	**1.00**	**.19**	**1.00**	**.17**	**1.00**	**.21**
Physical Appearance	**.19**	**1.00**	**.17**	**1.00**	**.21**	**1.00**
Peer Relationships	.25	.33	– –	– –	– –	– –
Opposite Sex	– –	– –	.23	.19	.23	.09
Same Sex	– –	– –	.13	.11	.12	.19
Honesty / Trustworthy	– –	– –	.11	.08	.15	.18
Parent Relationships	.08	.18	.05	-.03	.10	.09
Spiritual / Religions	– –	– –	– –	– –	-.01	-.01
Emotional Stability	– –	– –	.16	.18	.15	.18
Read / Verbal	.01	.04	.08	.08	.04	.13
Math	.07	.08	.08	.07	.09	.07
School	.10	.19	.14	.18	.06	.10
Problem solving	– –	– –	– –	– –	.11	.20
Global	.23	.39	.28	.31	.15	.33

Note: Physical Appr = Physical Appearance. The variables and numbers in bold show the intercorrelations between the two physical self-concept scales. Correlations are based on factor analysis results presented in manuals for the SDQI, SDQII, and SDQIII respectively.

Global Esteem and social components of self-concept, they are nearly unrelated to academic and other components of self-concept. It is also important to note that Physical Ability and Physical Appearance Self-concepts are not highly related to each other (*r*s about .2 for the three SDQ instruments). Because these two components of physical self-concept are so distinct, they should be considered separately and should not be incorporated into a single physical self-concept score that would confound the two components.

Self-concepts of very young children: An individual interview approach

Developing children's self-concepts is a critical educational goal in Australia and throughout the world. Despite considerable advances in self-concept theory, measurement, research, and practice with older students, there has been only limited progress with very young children 5-8 years of age. This is unfortunate as this developmental period may be crucial in the formation of a positive self-concept that is related to the attainment of many academic, social, physical, emotional, and developmental outcomes. This failure to pursue research with this very young age group is due, in large part, to problems associated with measuring self-concepts of very young children. In this study, we report preliminary findings based on 2 waves of self-concept and academic achievement data collected from children 5-8 years of age.

Marsh, Craven, and Debus (1991, 1998) described a new, adaptive procedure for assessing multiple dimensions of self-concept for children aged 5-8 using the SDQI. In considering this issue, we explored pictorial self-concept instruments (e.g., Harter & Pike, 1984), but found that the juxtaposition of the pictures and verbal explanations seemed more confusing to young students than the verbal presentations alone. In an individual interview format, the 64 positively worded items from the SDQI were administered to 501 kindergarten, 1st and 2nd grade students. The critical component was the individualised interview format used to collect SDQI responses. Procedures for the administration of the standard SDQI were adjusted to enable the modified SDQI to be administered as an individual interview and are described in greater detail by Marsh, Craven, and Debus (1991). The individual adminstration procedure began with instructions and four example items. After reading each example item, the interviewer asked the child if he/she understood the sentence. If the child did not understand the sentence the interviewer explained the sentence further, paraphrasing any words the child did not understand, ascertained if the child understood the sentence, re-read the sentence, and requested a response. In a strategy adapted in part from Harter and Pike (1984), the interviewer initially asked the child to respond 'yes' or 'no' to the sentence to indicate whether the sentence was true or false as a description of the child. If the child initially responded 'yes', the interviewer then asked the child if he/she meant 'yes always' or 'yes sometimes'. If the child initially responded 'no' the interviewer then asked the child if he/she meant 'no always' or 'no sometimes'. The second response probe was stated for every response even when it was answered in the initial response (e.g., the child said 'yes always' instead of 'yes'), thus providing a check on the accuracy of the child's initial response. After the child successfully responded to example items and any questions were answered, the interviewers then read aloud each of the 64 positively worded SDQI items. The child was encouraged to seek clarification of any item they did not understand. If the child stated that the item was not understood the interviewer explained the meaning of the item further and ascertained if the child understood the sentence before readministering the item. If the child indicated he/she understood the sentence but could not decide whether to respond yes or no, the interviewer recorded a response of 3, halfway between the responses of 'no sometimes' and 'yes sometimes'. Because this occurred infrequently and children were not told of this option, this middle category was seldom used. Halfway through the administration of the SDQI items the interviewer asked the child to do some physical activities for a brief period before proceeding to administer the remaining 32 items. This procedure was included to cater for young children's short attention spans.

There was an initial concern that the 64-item SDQI instrument would be too long for these very young children. Interestingly, items near the end were more effective than earlier items (in contrast to anticipated fatigue effects). Apparently, children learned to respond appropriately so that responses at the end of the instrument had much stronger psychometric properties than items at the beginning of the instrument. This observation has important implications for the typically short instruments used with young children. Based on confirmatory factor analyses (CFAs), Marsh et al. (1991) found support for all 8 SDQI scales, including the general self-concept scale, at each year level. However, with increasing age the differentiation among the 8 factors improved as inferred from the decreasing size of factor correlations. As part of this research, we compared their new assessment procedure with the standard group administration procedure in which the same SDQI items were read aloud to students. Kindergarten children were not able to complete this task, whereas the psychometric properties of group administration responses

were substantially poorer than those based on the individual interview responses for students in Years 1 and 2. In her review of self-concept instruments, Byrne (1996b) emphasised that the psychometric properties based on this instrument were stronger than those provided by any other instruments specifically designed for very young children.

More recently, Ellis, Marsh, and Craven (2000) extended this work to preschool children between the ages of 4 and 5.6 years. Preliminary work demonstrated that even with the individual interview format, the SDQI was not entirely suited for this preschool group. Based on extensive pilot research in which children were asked to explain their answers, some of the original SDQI items were modified or eliminated altogether and new items more suitable to this age group were developed. This resulted in a new 38-item Self-Description Questionnaire for preschool children (SDQP) that measures six self-concept factors (Physical, Appearance, Peers, Parents, Verbal, Math). The psychometric properties were good; the self-concept specific scales were reliable (.75 - .89; *Md* = .83), first and higher-order confirmatory factor models fit the data well, and correlations among the scales were moderate (-.03 - .73; *Md* = .29). Verbal and Math self-concepts, however, were much more highly correlated (.73) than found in previous research with older students. Physical self-concept was as highly correlated with the academic self-concept scales as the other nonacademic scales (including Appearance). In a higher-order factor analysis, three higher-order factors were identified (Academic; Physical; Nonacademic). Although not specifically directed at Physical self-concept *per se*, the results were very encouraging. The brief (6-item) Physical scale was well defined, the most reliable (.89) of the six scales, and clearly differentiated from the other six factors. Also, the original Physical self-concept scale from the SDQI instrument was more similar to the corresponding scale on the new SDQP than any other scales (5 of 6 items on the SDQP were from the SDQI).

Taken together, the results of these studies contribute to the critical debate in developmental psychology and early childhood research about the validity of self-reports for preschool children, demonstrating that self-reports by very young children do distinguish between multiple dimensions of self-concept at an even younger age than suggested by previous research.

Relations between SDQ Physical Self-concepts and external criteria

One approach to between-construct validation is to demonstrate that physical self-concept is substantially related to external criteria that are logically related to it. Nonphysical self-concepts should be less correlated to these external criteria. In support of these conclusions, results from two relevant studies (Marsh & Jackson, 1986; Marsh & Peart, 1988) summarized in Table 2, demonstrated that:

- SDQ Physical Ability Self-concept was significantly related to physical fitness, sport participation, physical activity levels, body mass index.
- SDQ Physical Appearance Self-concept was significantly related only to body mass index.
- SDQ Global Esteem was significantly related to only the body mass index.

Relations between SDQ Physical Self-concepts and known group differences

In the known group difference approach, groups "known" to differ on some characteristics related to physical self-concept are identified. If these groups differ substantially in physical self-concept, then there is support for the validity of physical self-concept. In an application of this approach (Marsh, Perry, Horsely, & Roche, 1995; also see Marsh & Jackson, 1986), elite Australian Institute of Sport (AIS) athletes were compared to

Table 2
Correlations between SDQ scales and external validity criteria

SDQ scales	Study 1		Study 2			
	Math Achieve (N = 137)	Physical Fitness (N = 137)	Body Mass (N = 45)	Physical Fitness (N = 45)	Sport Partic (N = 87)	Physical Activity (N = 87)
Physical Ability	**-.08**	**.45***	**-.36**	**.47***	**.39***	**.32***
Physical Appearance	**.13**	**.07**	**-.26***	**.17**	**.12**	**-.04**
Opposite Sex	-.16	.02	-.10	.29	.36*	.06
Same Sex	.09	.00	-.15	.23	.15	.19
Honesty / Trustworthy	.22*	-.07	-.20*	– –	– –	– –
Parent Relationships	.23*	-.09	-.16	– –	– –	– –
Emotional Stability	.12	.11	-.11	– –	– –	– –
Read / Verbal	.27*	.00	-.10	– –	– –	– –
Math	.48*	-.09	-.01	– –	– –	– –
School	.50*	-.08	-.17	– –	– –	– –
Global	.23*	.07	-.27	.14	-.03	.19

Note: Math Achieve = Math Achievement; Sport Partic = Sport Participation. The variables and numbers in bold show the correlations of the two physical self-concept scales with external criteria. Physical fitness is defined by a composite score based on a diverse set of field exercises in both Study 1 (400M run, push-ups, burpees, jump rope, sit-ups, v-sits, and step-ups) and 2 (sit-ups, bend-twist-touch test, sit-and-reach, burpee, step-test). In Study 2, participation is the number of sports participated in, whereas physical activity is the number of at least moderately strenuous activities the person participated in. Studies 1 and 2 are described in greater detail in Marsh and Peart (1988) and Marsh and Jackson (1986), respectively.

a large normative sample of nonathletes on the 13 SDQIII scales (see Table 3).

- Athletes had substantially higher Physical Ability Self-concepts, but did not differ on Appearance Self-concept.
- Athletes were somewhat higher on Social Self-concepts (Same Sex, Opposite Sex, Parent) and Esteem.
- Group differences were nonsignificant for Academic Self-concepts (Math, Verbal, Academic, Problem Solving) and Emotional Self-concept, whereas athletes had marginally lower Spiritual and Honesty Self-concepts.
- Athlete/nonathlete differences interacted with gender, mostly favoring women athletes.

There were large differences in Physical Ability Self-concept and smaller or nonsignificant differences in academic and nonphysical self-concepts. This pattern supports *a priori* predictions and the SDQIII construct validity. The specificity of known group differences to particular components of self-concept shows that self-concept cannot be understood if its multidimensionality is ignored.

Relations between SDQ Physical Self-concepts and experimental manipulations: Outward Bound Programs

Two Outward Bound interventions and their contrasting predictions provide a powerful test of

Table 3
Known group difference approach: Mean SDQIII scales for elite athletes (AIS) and normative comparison group

SDQ scales	AIS		Norm		Overall ANOVA		
	Male $N = 36$ Mean	Female $N = 47$ Mean	Male $N = 1489$ Mean	Female $N = 896$ Mean	Group t(2414)	Gender t(2414)	Gender X Group t(2414)
Physical Ability	**6.81**	**6.99**	**6.10**	**5.57**	**8.08**	**1.34**	**-2.66****
Physical Appearance	**5.48**	**4.50**	**5.49**	**4.73**	**-1.04**	**7.82****	**.99**
Math	5.03	5.16	5.31	4.79	.26	1.13	-1.85
Verbal	5.33	5.74	5.56	5.53	-.07	-1.64	-1.84
Academic	5.21	5.86	5.67	5.67	-1.06	-2.49*	-2.50*
Problem Solving	5.03	5.41	5.44	4.98	.08	.38	-3.95**
Same Sex	6.00	6.23	5.73	5.80	3.23**	-1.37	-.69
Opposite Sex	5.67	5.72	5.22	5.38	2.75**	-.71	.37
Parent	6.50	6.63	5.95	5.89	5.07**	-.28	-.80
Spiritual	4.25	4.08	4.36	4.96	-2.56**	-1.12	1.98*
Honesty	5.57	6.46	6.16	6.21	-2.06*	-5.69**	-5.06
Emotional	5.54	5.37	5.59	5.26	.21	1.96*	-.59
Esteem	6.43	6.09	6.06	5.73	2.95**	2.74**	.09
Total Self	5.60	5.73	5.59	5.42	2.28*	.28	-2.14*

Note. The number in bold point out the two Physical self-concept scales.Self-concept responses from 83 elite athletes (AIS) and 2436 participants from the normative archive for the SDQIII instrument (Norm Group) were compared using a 2 (Group: AIS or Norm) x 2 (Gender) analysis of variance. Separate tests were also conducted comparing group differences (AIS vs. Norm groups) separately for each gender, and comparing gender differences separately for each group. All means and standard deviations vary along an 8-point response scale used for the SDQ (1 = lowest, 4.5 = middle, 8 = highest). For a more detailed description of this research see Marsh, Perry, Horsely, & Roche (1995) from which this table was derived. * $p < .05$; ** $p < .01$.

the multidimensionality of self-concept.

• The Outward Bound Standard Course is a 26-day residential program of physical outdoor activities. Program goals were primarily nonacademic. It was predicted and found that the program affected primarily the physical and other nonacademic self-concepts. It had little impact on academic self-concept. The size and pattern of effects were maintained in an 18 month follow-up (Marsh, Richards, & Barnes, 1986a, 1986b).

• The Outward Bound Bridging Course was developed for low-achieving high school males to improve math and reading. Program goals were primarily academic. It was predicted and found that the program primarily affected academic self-concepts (and reading and math achievement). It had little effect on nonacademic self-concepts (Marsh & Richards, 1988a).

Support for both these contrasting sets of predictions provides particularly strong support for the use of multidimensional self-concept measures in intervention studies. In both studies, it was also argued that the inclusion of less relevant self-concept scales provided a test for halo effects and placebo-like biases. Hence, the close match between the intent of the intervention and a pattern

of results for multiple dimensions of self-concept provides an important support for the construct validity of interpretations of both the intervention itself and the multidimensional self-concept measures used to evaluate the intervention.

Relations between SDQ Physical Self-concepts and experimental manipulations: Competitive/Cooperative Aerobics Intervention

Marsh and Peart (1988) randomly assigned high school students to competitive, cooperative and control groups.

- The cooperative group completed exercises in pairs; feedback emphasized individual improvement.
- The competitive/social comparison group completed individual exercises; feedback emphasized comparisons with whoever did best on each exercise.

Consistent with *a priori* predictions Marsh and Peart (1988) found that:

- The cooperative intervention increased physical fitness and physical self-concept.
- The competitive intervention increased physical fitness but decreased physical self-concept.
- Other self-concepts were unaffected.

Critical features were: frame of reference effects and social comparison processes. Students in the competitive group knew their fitness had improved. However, they were forced to compare their performances with whoever did best on each exercise. Thus, their frame of reference used to evaluate their performances changed even more than did their fitness levels. The net effect of the competitive intervention on Physical self-concept was negative. In highly competitive environments, there are likely to be many "losers" and few "winners" and this is likely to lead to lower levels of self-concept.

The research also demonstrates why it is important to assess physical self-concept even when the focus is on skill development or fitness enhancement. Without the inclusion of a physical self-concept measure, the competitive intervention would have been evaluated positively – just as positive as the cooperative intervention in terms of short-term fitness enhancement. Only the inclusion of the PSDQ demonstrated that there were unintended negative effects associated with the competitive intervention that would likely undermine any long-term gains associated with it. Short-term gains are more likely to be maintained if there is an increase in self-concept. If interventions inadvertently undermine self-concept, short-term gains are unlikely to be maintained. More generally, as shown in the academic area, physical skill self-concept and physical attributes such as physical skill development and fitness are likely to have reciprocal effects. The best way to enhance and maintain development in either one is to enhance both.

Physical self-concept: Relations with physical fitness for boys and girls

Although not based on SDQ responses, Marsh (1993c) used data from the Australian Health and Fitness survey to relate academic and physical self-concepts to a diverse set of physical fitness indicators and to academic achievement. Participants (aged 9-15) rated their physical fitness, health, and academic achievement. They also completed an extensive battery of field and technical indicators of fitness. Consistent with predictions (see Table 4) Marsh (1993c) found that:

- Physical and Academic Self-concepts were distinct and became more distinct with age (low correlations and contrasting relations with physical fitness and academic achievement).
- Validity coefficients relating Physical and Academic Self-concept to objective measures increased with age.
- Relation between fitness and Physical self-

Table 4
Relations between multiple dimensions of self-concept and measures of physical fitness and academic achievement

		Self-concepts of		
	N	Fitness	Health	Academic
		Fitness indicators		
1.6K Run	5840	.354*	.170*	.052*
50M Dash	5975	-.281*	-.116*	-.045*
Long Jump	6258	.256*	.118*	.057*
Push Up	6239	.292*	.149*	.009
Sit Up	6266	.207*	.102*	.023
Sit & Reach	6260	.042*	.047*	.040*
Body Mass	6284	-.225*	-.132*	-.009
Body Girths	6281	-.010*	-.124*	.000
Static Strength	2655	.040	-.001	.010
Skinfold	2681	-.329*	-.180*	-.010
Blood Press	2649	-.078*	-.071*	-.003
PWC170	2563	.287*	.127*	.019
VO2max	270	.295*	.252*	.103
Lung Capacity	2669	.008	-.005	.032
		Academic achievement		
School Work	5890	.088*	.131*	.418*
		Self-concepts		
Physical Fitness	6283	– –		
Health	6286	.345*	– –	
Academic	6277	.141*	.163*	– –

Note. Data used in this study is from the Australian Health and Fitness Survey. For a more detailed description of this research see Marsh (1993c) from which this table was derived.
* $p < .01$

concept was similar for boys and girls.

- Consistent with a multidimensional perspective many different components of physical fitness contributed to Physical Fitness Self-concept.

Commenting on limitations of the study, Marsh specifically noted that physical fitness self-concept can be divided into subcomponents in the same way that general academic self-concept is usefully subdivided into self-concepts associated with specific school subjects. Such a multifaceted, hierarchical structure of physical self-concept is clearly consistent with the Marsh/Shavelson model of self-concept and particularly their more differentiated, multidimensional model of academic self-concept. Marsh proposed that a particularly fruitful direction for such research would be to relate a

multidimensional profile of physical fitness indicators to a parallel set of multidimensional physical self-concept scales.

Physical SDQ (PSDQ) Instrument

Historical background

Wylie (1974, 1989) evaluated a wide variety of self-concept measures. Her 1974 review revealed that at the time most self-concept instruments focused on global self-concept or self-esteem rather than specific domains such as physical self-concept. Although several of the instruments reviewed by Shavelson et al. (1976) contained items relating to physical skills and elements of physical appearance, none provided a clearly interpretable measure of physical self-concept. From a practical perspective, these older instruments appear to be of little value for sport and exercise psychologists. The major exception to this conclusion was, perhaps, the Physical Estimation and Attraction Scales (Sonstroem, 1978, 1988) instrument and the theoretical model on which it was based. This instrument was designed to measure two global components: Estimation (competency) and Attraction, and has had considerable impact on sport psychology research.

In her more recent review, Wylie (1989) identified several multidimensional self-concept instruments that measure one or more components of physical self-concept that can be differentiated from other specific domains of self-concept and general self-concept. Included in her list were the set of three SDQ instruments already discussed. Wylie also evaluated Harter's (1985) Self-Perception Profile for Children that contains two physical self-concept scales (Athletic Competence and Physical Appearance). Other multidimensional instruments that contain physical scales that were not reviewed by Wylie include the Self-rating Scale (Fleming & Courtney, 1984) that measures physical ability and physical appearance, the Song and Hattie test (Hattie, 1992) that measures physical appearance, and the Multidimensional Self-Concept Scale (Bracken, 1992) that has a physical scale which includes physical competence, physical appearance, physical fitness, and health. The Tennessee Self-Concept Scale (Fitts, 1964) is a multidimensional self-concept instrument that also purported to measure physical self-concept. In their review and empirical evaluation of this instrument, Marsh and Richards (1988b) found distinguishable physical components reflecting health, neat appearance, physical attractiveness, and physical fitness that were incorporated into a single physical self-concept score. This more detailed breakdown of the Tennessee physical scale was supported by relations with the SDQ Physical Ability and Physical Appearance scales in a MTMM study comparing responses to the two instruments. Because each of these clusters based on responses to the Tennessee instrument was represented by only a few items, however, it would not be appropriate to use the instrument to measure these distinct components of physical self-concept. Marsh and Richards argued that physical scores – like those based on the Tennessee Self-Concept Scale – that combine and confound such a wide range of differentiable physical components should be interpreted cautiously (also see similar comments by Fox & Corbin, 1989).

In summary, from a historical perspective, most self-concept instruments have either ignored physical self-concept completely or have treated physical self-concept as a relatively unidimensional domain incorporating characteristics as diverse as fitness, health, appearance, grooming, sporting competence, body image, sexuality, and physical activity into a single score. This concern led me to develop the Physical Self-Description Questionnaire (PSDQ).

PSDQ psychometric properties

The theoretical basis and design of the PSDQ

Table 5
Physical Self Description Questionnaire (PSDQ): Summary of the content of the nine specific and two global scales

Scale	Summary
Appearance:	Being good looking; having a nice face.
Strength:	Being strong; a powerful body; lots of muscles.
Condition/Endurance:	Being able to run a long way without stopping; not tiring easily when exercising hard.
Flexibility:	Being able to bend and turn you body easily in different directions.
Health:	Not getting sick often; getting well quickly.
Coordination:	Being good at coordinated movements; being able to do physical movements smoothly.
Activity:	Being physically active; doing lots of physical activities regularly.
Body Fat:	Not being overweight; not being too fat.
Sport:	Being good at sports; being athletic; having good sports skills.
Global Physical:	Feeling positive about ones physical self.
Global Esteem:	Overall positive feelings about self.

follows SDQ research. PSDQ scales reflect some SDQ scales (Physical Ability, Physical Appearance, and Esteem). They also reflect my attempt to parallel physical fitness components identified in my confirmatory factor analysis of physical fitness (Marsh, 1993b), extending Fleishman's (1964) classic research on the structure of physical fitness. Each PSDQ item is a simple declarative statement and individuals respond using a 6-point true-false response scale (like on SDQII). The PSDQ is designed for adolescents, but should be appropriate for older participants. The content of the PSDQ scales is summarized in Table 5 (also see Appendix 1) and psychometric properties are summarized in Table 6. These results demonstrate:

- Good reliability (Median coefficient alpha = .92 across the 11 scales (Marsh, 1996b; Marsh, Richards, et al., 1994; see Table 6).
- Good test-retest stability over short-term (median r = .83 for 11 PSDQ scales, 3 months) and longer-term (median r = .69, 14 months (Marsh, 1996b; see Table 6).
- A well defined, replicable factor structure as shown by confirmatory factor analysis (Marsh, 1996b; Marsh, Richards, Johnson, Roche, & Tremayne, 1994).
- A factor structure that is invariant over gender as shown by multiple group CFA (Marsh, Richards, et al., 1994).
- Convergent and discriminant validity as shown by a MTMM study of responses to three Physical Self-concept instruments (see Marsh, Richards, et al., 1994, and subsequent discussion of Table 7).
- *Convergent and discriminant validity* as shown by PSDQ relations with external criteria (see Marsh, 1996a, and subsequent discussion of Table 8).
- Applicability for participants aged 12 to 18 (or older) and for elite athletes and nonathletes.

The PSDQ is a psychometrically strong instrument that is appropriate for a wide variety of sport/exercise research.

Table 6
Reliability and stability for the 11 PSDQ scales on four times (T1 - T4)

	Reliability coefficients				Stability coefficients					
	T1	T2	T3	T4	T1/T2 7mo	T1/T3 10mo	T1/T4 14mo	T2/T3 3mo	T2/T4 7mo	T3/T4 4mo
PSDQ scale										
Health	.87	.82	.82	.82	.68	.67	.31	.70	.40	.42
Coordination	.91	.90	.93	.90	.79	.68	.69	.83	.74	.78
Physical Activity	.90	.91	.95	.91	.74	.67	.67	.82	.77	.76
Body Fat	.96	.96	.96	.95	.87	.84	.79	.89	.80	.84
Sport	.94	.95	.95	.94	.87	.79	.82	.87	.84	.86
Global Physical	.96	.95	.96	.95	.78	.73	.72	.82	.76	.82
Appearance	.91	.92	.94	.93	.73	.66	.67	.78	.70	.77
Strength	.92	.93	.92	.91	.81	.79	.75	.89	.74	.76
Flexibility	.90	.90	.92	.90	.72	.68	.67	.83	.76	.78
Endurance	.92	.92	.95	.93	.82	.82	.78	.87	.85	.86
Global Esteem	.91	.92	.94	.92	.80	.77	.66	.89	.75	.76
Median	.91	.92	.94	.92	.79	.73	.69	.83	.76	.78

Note. T1 - T4 represent four testing occasions; mo = number of months separating each pair of testing occasions (for purposes of assessing stability).
For a more detailed description of this research see Marsh (1996b) from which this table was derived.

Multidimensional, hierarchical physical self-concept and multiple components of physical fitness

Marsh and Redmayne (1994) related 6 Physical self-concepts [Endurance, Balance, Flexibility, Strength, Appearance, and Global Physical Ability] to five physical fitness tests [Endurance, Balance, Flexibility, Static Strength, Explosive Strength/Power] for girls (aged 13/14). The hierarchical confirmatory factor analyses:

- identified the six Physical Self-concept scales on this early version of the PSDQ
- provided support for a multidimensional, hierarchical model of physical self-concept.

The pattern of correlations between specific components of physical self-concept and physical fitness generally supported the construct validity of the PSDQ responses.

The correlation between global Physical Self-concept and general physical fitness (r = .76) was substantial.

Multitrait multimethod comparison of three physical self-concept instruments

The MTMM design is used to test convergent, discriminant, and construct validity (see Marsh, 1988, for a general discussion of the MTMM design and the analysis of MTMM results). Reviewers of self-concept measurement (e.g., Byrne, 1984, 1996b; Hattie, 1992; Marsh, 1990b; Shavelson et al., 1976; Wylie, 1974; 1979; 1989) emphasize the central role of MTMM analyses in the construct validation of self-concept responses. In this approach, multidimensional self-concept instruments purporting to measure the

same or substantially overlapping scales are administered to the same group of respondents. The approach consists of a systematic evaluation of correlations between scales from different instruments that are posited to be matching (the same or similar content) and nonmatching. In this approach, convergent validity is supported by large correlations between matching scales from different instruments and discriminant validity is supported when convergent validities are larger than other correlations.

MTMM analyses also reveal important problems in the interpretation scale scores based on the label that is attached to them by their author or other researchers (e.g., Marsh, 1994). Thus, for example, the Jingle Fallacy (Marsh, 1994) is assuming that two scales with the same label measure the same construct and the Jangle Fallacy is assuming that two scales with different labels measure different constructs. Given the prevalence of the MTMM design in self-concept research and, more generally, in most areas of psychological measurement, it is surprising that the technique has not been used more widely in sport and exercise research.

The Marsh, Richards, Johnson, Roche, and Tremayne (1994) MTMM study is important because it is apparently the first MTMM study of multidimensional physical self-concept instruments, because it reveals some important concerns in the three instruments, and because it provides a possible model for other physical self-concept research. In this study students completed three physical self-concept instruments: the PSDQ and two other multidimensional physical self-concept instruments designed to measure scales that overlap with those from the PSDQ:

- Fox's Physical Self-Perception Profile (PSPP; Fox, 1990; Fox & Corbin, 1989) is widely used. It has 5 scales: Physical Condition, Physical Strength, Body Attractiveness, Sport, Physical Self Worth.
- Richard's Physical Self-Concept (PSC; Richards, 1987, 1988) has been used mostly in Outward Bound research. It has 7 scales: Activity, Appearance, Health, Competence, Strength, Body Build, Satisfaction. Its 7-factor solution is remarkably robust over gender and age (10 through 60).

We first compared the content of items from the 23 scales (11 PSDQ, 5 PSPP, 7 PSC). We predicted which scales from different instruments should be most correlated – the convergent validities. The results supported these predictions in that convergent validities were consistently large and larger than correlations among nonmatching factors (Table 7).

Based on these results we concluded that:

- The PSDQ Strength, Sport, Physical Activity, Coordination, Endurance, Health, Physical Appearance, and Global Physical scales were substantially correlated with corresponding scales from the other two instruments. PSDQ convergent validities were higher than those involving the other two instruments.
- PSDQ Body Fat was distinct from, but related to, the Body scales on other two instruments.
- Flexibility and Global Esteem scales were not included on the other two instruments.
- Correlations (convergent validities) among scales from the three instruments predicted to be matching were systematically larger than those among nonmatching scales.

These results support the convergent and discriminant validity of PSDQ responses. The comparison of responses from the PSDQ, PSPP, and PSC instruments (see Marsh, Richards, et al., 1994, for more detailed discussion) provided a demanding test of the construct validity of the three physical self-concept instruments. Overall, the results provided good support for the convergent and discriminant validity of responses to the three instruments.

It is also of practical importance to evaluate these results in terms of the relative usefulness of the three physical self-concept instruments. Psychometrically, the PSPP appeared to be the weakest of the three instruments for responses

Table 7
Correlations among 23 scales from 3 physical self-concept instruments

Scales	PSDQ								PSPP												PSC		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
PSDQ (Marsh)																							
1 STRG	1																						
2 BFAT	05	1																					
3 PACT	56	24	1																				
4 ENDR	56	39	68	1																			
5 SPRT	64	36	76	72	1																		
6 CORD	56	46	70	57	78	1																	
7 HEAL	21	27	24	18	25	37	1																
8 APPR	39	45	34	38	50	58	14	1															
9 FLEX	42	36	51	58	47	62	31	34	1														
10 GPSC	50	60	57	50	65	77	36	64	50	1													
11 ESTM	47	56	49	50	60	66	58	60	56	74	1												
PSPP (Fox)																							
12 STRG	**86**[a]	20	52	51	62	50	23	39	34	47	47	1											
13 BODY	42	**61**[b]	36	46	49	53	14	**68**[b]	37	67	51	65	1										
14 COND	49	45	**73**[b]	**70**[b]	68	58	20	38	45	60	51	67	71	1									
15 SPRT	58	38	69	69	**86**[a]	67	25	47	45	59	55	69	70	89	1								
16 GPSW	54	51	57	60	64	62	24	54	45	**81**[a]	68	73	82	85	86	1							
PSC (Richards)																							
17 STRG	**90**[a]	02	48	43	56	47	16	36	31	42	40	**79**[a]	41	40	51	50	1						
18 BODY	46	**68**[b]	46	53	58	60	22	65	44	74	63	46	**81**[a]	60	61	70	46	1					
19 PACT	44	23	**66**[b]	53	62	54	36	18	33	37	45	40	22	58	63	48	38	35	1				
20 COMP	52	33	57	56	69	**84**[a]	32	57	59	59	62	51	49	50	70	65	59	64	58	1			
21 HEAL	25	10	25	22	28	37	**83**[a]	18	23	32	42	21	12	20	31	22	15	23	41	41	1		
22 APPR	34	33	34	29	43	52	18	**88**[a]	25	58	51	33	61	33	41	49	41	64	15	54	21	1	
23PSAT	28	38	32	38	41	38	20	33	30	39	38	36	39	41	55	42	31	42	32	44	14	26	1

Note. STRG = strength, BFAT = body fat, PACT = physical activity, ENDR = endurance/fitness, SPRT = sports competence, CORD = coordination, HEAL = health, APPR = appearance, FLEX = flexibility, GPSC = general physical self-concept, ESTM = esteem, COND = condition, GPSW = general physical self worth, COMP = competence, PSAT = physical satisfaction. All coefficients vary between 0 and 1 (decimal points are not presented to conserve space). Scales from different instruments that are predicted to be most highly correlated (i.e., the convergent validities in MTMM analysis) are in bold and underlined. Factor correlations were based on a confirmatory factor analysis of responses to all three physical self-concept instruments in which each variable was allowed to load only the factor that it was designed to measure and all other factor loadings were constrained to be zero (see Marsh, Richards, et al., 1994, for factor loadings and further analyses).

[a] Convergent validities between scales predicted a priori to be most closely matched. [b] Convergent validities between scales predicted a priori to be less closely matched.

by Australian high school students (as opposed, perhaps, to US university students for whom the instrument was designed). The coefficient alpha estimates of reliability were systematically lower for the PSPP than for either of the other instruments. The CFA analyses consistently demonstrated that the PSPP responses had lower trait factor loadings and more measurement error, whereas there was also evidence suggesting a systematic method effect apparently associated with the nonstandard response scale used on the PSPP. The very large correlations among the PSPP factors seemed to undermine support for the instrument's ability to differentiate among the factors that it was designed to measure. Furthermore, two of the PSPP scales seem to combine potentially important components of physical self-concept that are measured with separate scales on the PSDQ (the PSPP Condition scale with the PSDQ Physical Activity and Physical Fitness/Endurance scales and, perhaps, the PSPP Body with PSDQ Body Fat and Appearance scales). Also, Marsh, Richards, et al. (1994) reported that some subjects had difficulty completing the nonstandard PSPP response scale, a finding that was consistent with the Marsh and Gouvernet (1989) and Marsh and McDonald-Holmes (1990) studies of the original Harter instrument that was the basis of the PSPP response format. It is, however, likely that this difficulty could be overcome with more detailed instructions and closer monitoring of the completion of the instrument. For these reasons, Marsh, Richards, et al. recommended that the PSPP responses by young adolescents should be interpreted cautiously and that these concerns should be pursued in a replication of their MTMM study with older subjects and university students for whom the PSPP was designed.

The comparison of the PSDQ and PSC instruments is not so straightforward. The PSDQ is a more comprehensive instrument in that it measures a much broader range of physical self-concept components. This was not, however, the intended purpose of the PSC which was designed to provide a quick, reliable measure of a limited number of components of physical self-concept that were widely applicable across gender and age. Whereas the psychometric properties of the PSDQ appeared to be slightly stronger than those of the PSC, the differences were not substantial. Hence, the major differences seem to be the brevity of the PSC compared to the comprehensiveness of the PSDQ. (Depending on the age of the subjects, the PSC can be completed in 5-10 minutes whereas it takes 10-15 minutes to complete the PSDQ.) It is also relevant, however, to recommend care in the interpretation of some of the PSC scales. The PSC Physical Competence scale should probably be interpreted as a self-concept measure of physical coordination and, perhaps, agility. Also, the PSC Satisfaction scale should not be interpreted as a measure of general or global physical self-concept, as evidenced by the modest correlations between this scale and the general physical scales from the PSDQ and PSPP instruments. Indeed, the PSC Satisfaction scale appears to reflect a complicated combination of the importance placed on a particular component of physical self-concept and how well one matches one's ideals. High satisfaction scores could reflect high levels of accomplishment (I am satisfied because I am physically competent), low ideals (I am satisfied because my expectations are low), or a lack of importance placed on physical competence (I am satisfied because I don't really care whether I am physically competent). Whereas the PSC Satisfaction scale measures a potentially important aspect of physical self-concept, responses to this scale are likely to be complicated and should be interpreted appropriately.

Marsh, Richards, et al. (1994) also noted potential limitations in their study that dictated some caution in the interpretation of the PSDQ responses and provided directions for additional research. In particular, because the PSPP was designed for use by American university

students, it may be unfair to evaluate it on the basis of responses by Australian high school students. It would, however, be useful to replicate this study with American university students – as well as students of different ages from different countries – in order to test the generalizability of the findings. Also, because this study was the first application of the 11-scale PSDQ instrument, there is clearly need for further research to further evaluate the PSDQ responses in relation to external criteria, although similar concerns can be expressed in relation to the other two instruments as well. Marsh, Richards, et al. suggested that the PSDQ was appropriate for older subjects and for US students (like those targeted by the PSPP), but it is important to test the generalizability of the PSDQ's psychometric properties with different populations.

Recent research (Marsh, Asci, & Tomas, 2002; Marsh, Marco, & Asci, in press) extended this research in two studies to evaluate the cross-cultural construct validity of physical self-concept responses. In Marsh, Marco, and Asci (in press) we showed that the factor structure of the Physical Self Descriptionaire (PSDQ) was reasonably invariant over large samples of responses by Australian, Spanish, and Turkish students. Although there was reasonable support for the complete invariance of factor loadings, factor correlations and factor variances across all three groups, the factor structures based on Austalian and Spanish high school students were somewhat more similar to each other than to the factor structure based on Turkish university students. In Marsh, Asci, and Marco (2002) we demonstrated support for convergent and discriminant validity of PSDQ responses in a multitrait-multimethod analysis of relations with responses to the Physical Self Perception Profile (PSPP) based on responses by Turkish university students. In support of construct validity interpretations, matching PSDQ and PSPP factors were highly correlated. However, support for the PSPP was undermined by extremely high correlations among several of its factors, apparently due in part to a substantial method effect associated with its idiosyncratic response scale. Based on psychometric, theoretical, cross-cultural, and practical considerations, the results support the use of the PSDQ in a wide variety of research and applied settings.

A unique, new aspect of the study was the focus on the systematic evaluation of physical self-concept responses in different cultures. Particularly in Marsh, Marco, and Asci (in press) the cross-cultural results provided strong support for the appropriateness of the PSDQ instrument for Spanish high school students and Turkish university students as well as the Australian high school students for whom it was originally developed. These results address concerns about the appropriateness of the PSDQ for older, university-aged respondents expressed by Marsh and colleagues (Marsh, 1997; Marsh, Richards, et al., 1994) and the more general concern about the appropriateness of the various SDQ instruments for non-Australian settings (Keith & Bracken, 1996). More generally, Marsh, Marco, and Asci (in press) addressed technical details in cross-cultural comparisons that are typically not addressed in sport and exercise research. In Marsh, Asci, and Marco (2002) we replicated and extended aspects of the Marsh, Richards, et al. (1994) MTMM study. Importantly, the results of Study 2 were based on responses by Turkish university students to translated versions of two of the most widely used physical self-concept instruments (the PSDQ and the PSPP). As in the original research, the results of this new MTMM study provided good support for the construct validity of physical self-concept as well as several of its underlying elements. This type of detailed analysis provides unique insight into the intricacies of item and scale design and content and the way wording is interpreted by the population under scrutiny.

Relations to external criteria

Marsh (1996a) related PSDQ responses to 23

Table 8
Relations between PSDQ scales and external validity criteria

Variable	BFAT	APPR	PACT	ENDR	STRG	FLEX	CORD	SPORT	HEAL	GPHY	ESTM
Body Composition											
Body Mass Index	-.69	-.19	-.17	-.32	.13	-.29	-.25	-.21	-.00	-.32	-.16
Girths	-.70	-.21	-.17	-.34	.14	-.33	-.27	-.23	.01	-.31	-.14
Skinfolds	-.66	-.23	-.18	-.41	-.10	-.31	-.29	-.37	.00	-.41	-.17
Body Composite	-.72	-.23	-.20	-.38	.04	-.32	-.30	-.29	-.01	-.37	-.17
Silhouette Ratings											
Actual	-.66	-.22	-.17	-.39	-.06	-.40	-.38	-.34	-.07	-.43	-.30
Actual-Ideal	-.76	-.27	-.20	-.40	-.00	-.36	-.36	-.34	-.01	-.48	-.35
Program Participation											
Past Participation	.21	.20	.48	.42	.27	.32	.27	.40	.20	.35	.34
Present Participation	.12	.26	.59	.49	.33	.36	.34	.41	.08	.33	.38
Future Desire	.04	.23	.50	.32	.24	.30	.35	.38	.09	.26	.26
Future Intention	.03	.15	.54	.37	.33	.29	.36	.41	.03	.26	.27
Barriers	-.20	-.30	-.54	-.49	-.41	-.45	-.50	-.55	-.17	-.47	-.57
Benefits	-.21	-.03	.24	.09	.14	.07	.09	.10	.00	-.06	.00
Commitment	-.09	.07	.27	.16	.23	.14	.25	.18	.11	.07	.09
Activity levels											
Times/typical week	.06	.14	.55	.48	.40	.32	.38	.39	.08	.26	.32
Hours/typical week	.13	.16	.37	.36	.29	.28	.32	.30	.17	.24	.25
Hour last 7 days	.14	.22	.36	.39	.30	.26	.31	.41	.17	.26	.25
Endurance											
1.6km Run	-.51	-.22	-.38	-.63	-.35	-.37	-.40	-.54	.06	-.48	-.33
Beep Test	.42	.14	.35	.54	.25	.34	.35	.48	.02	.44	.34
Dynamic/Explosive Strength											
Basketball Throw	.04	.01	.09	.20	.40	.15	.17	.27	.04	.21	.23
Modified Pull Up	.35	.15	.22	.38	.30	.30	.29	.32	-.01	.28	.26
50m Dash	-.32	-.08	-.20	-.35	-.23	-.21	-.23	-.29	.03	-.24	-.19
Stand Long Jump	.21	.15	.19	.33	.38	.26	.35	.37	.01	.24	.27
Flexibility											
Sit and Reach	.11	.05	.15	.12	.01	.37	.25	.13	-.13	.07	.08

Note. STRG = strength, BFAT = body fat, PACT = physical activity, ENDR = endurance/fitness, SPORT = sports competence, CORD = coordination, HEAL = health, APPR = appearance, FLEX = flexibility, GPHY = general physical self-concept, ESTM = esteem. For a more detailed description of this research see Marsh (1996a) from which this table was derived.

external validity criteria: measures of body composition, physical activity, endurance, strength, and flexibility. Each criterion was predicted to be most highly correlated to one of the PSDQ scales (see Table 8).

- In support of convergent validity, every predicted correlation was significant.
- In support of discriminant validity, most predicted correlations were larger than other correlations involving the same criterion.

This pattern of relations between PSDQ responses and external validity criteria support the construct validity of PSDQ responses.

PSDQ Body Fat and Appearance

Now I focus specifically on relations among PSDQ Body Fat, PSDQ Appearance, objective body composition, and silhouette ratings. In particular, I want to test the need for separate PSDQ Body Fat and PSDQ Appearance Self-concepts that is a unique feature of the PSDQ.

Objective body composition. The objective body composition score was based on body mass index, multiple measures of body girths, and skinfolds at multiple sites (see Table 8).

- Objective body composition is substantially related to PSDQ (lack of) Body Fat (r = -.72)
- Objective body composition is much less correlated with PSDQ Physical Appearance (r = -.23).
- Objective body composition correlated more highly with many PSDQ scales than PSDQ Appearance (e.g., -.38 with PSDQ Endurance, -.32 with PSDQ Flexibility; -.30 with PSDQ Coordination; -.29 with PSDQ Sport), but none of these correlations approached the size of correlations between objective body composition and PSDQ Body Fat.

These results support the need for separate PSDQ Body Fat and PSDQ Physical Appearance scales. Objective measures of body composition are substantially related to PSDQ Body Fat and are not substantially related to PSDQ Physical Appearance.

Discrepancy theory: Actual and ideal body image. As noted earlier, William James (1890) emphasized 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," leading him to conclude that objective accomplishments are evaluated in relation to internal frames of reference. Following from James and others, discrepancy theory posits self-concept is a function of differences between self-perceived actual accomplishments and ideal standards so that:

- similar accomplishments lead to different self-evaluations, depending on ideal standards;
- unrealistic ideals lead to poor self-concepts even when accomplishments are otherwise good.

Despite a history of criticism and limited empirical support, this discrepancy model has led to a century of heuristic speculation, empirical research, theoretical debate, and conflicting claims.

Marsh and Roche (1996) devised a new test of discrepancy theory based on silhouette ratings. Students selected one of 9 silhouettes varying on a ectomorphy-endomorphy continuum to represent their actual and ideal body image. Results summarized in Table 8 indicate that:

- Silhouette ratings were substantially correlated (r = .62) with objective body composition (BMIs, girths, and skinfolds)
- PSDQ Body Fat was substantially correlated to Actual silhouette ratings (r = .66), but even more correlated with Actual-Ideal discrepancies (r = .76).
- Actual ratings contribute positively to the prediction of self-concept, but high Ideal ratings contribute negatively.
- Similar (but smaller) patterns of relations were evident for Global Physical Self-concept and Esteem.

More complicated models developed by Marsh and Roche (1996) showed that taking into

account the direction as well as the size of the actual-ideal discrepancies did even better. The results support Discrepancy Theory. Low self-concepts may reflect poor Actual self-perceptions and/or unrealistic ideal standards.

Marsh (1999) extended the earlier Marsh and Roche (1996) study by developing a new, expanded Silhouette rating scale and evaluating the role of "future self" and "potential self" as well as "ideal self." Participants (793 high school students) indicated their Actual, Ideal, Future, and Potential body image by selecting from 12 silhouettes varying along an obese-skinny continuum and completed 7 self-concept factors on two occasions. Consistent with a priori predictions, structural equation models demonstrated that actual-self had positive effects on self-concept factors whereas the effects of ideal-self were negative (more demanding, increasingly slender ideals detracted from self-concept), making this one of the few studies to support discrepancy models' predictions using this paradigm. Furthermore, Actual and Ideal effect sizes varied systematically for self-concepts selected to be more or less related to body image. There was, however, little support for Future-self and Potential-self factors as standards against which to evaluate actual self. The effects of Future-self on self-concept after controlling for Actual-self tended to be positive, not negative. Thus, an optimistic self-perception of one's future body image tended to have a positive effect on self-concept and may not provide a negative basis of comparison for evaluating one's current body image as implied by cognitive discrepancy models. There was some support for Potential-self in that predictions based on Actual and Potential selves were better than those based on Actual-self alone. However, Potential and Ideal selves were very highly correlated, the combination of Potential, Actual, and Ideal selves did little better than Actual and Ideal selves, and the contribution of Potential self after controlling Ideal self tended to be positive, not negative.

Implicit in Discrepancy theory is the untested assumption that larger actual-ideal discrepancies lead to lower self-concepts. However, causal ordering cannot be adequately evaluated with a single wave of data. Marsh (1999) extended previous research by evaluating structural equation models of longitudinal data in which the same Silhouette and self-concept ratings were collected on two occasions. This multiwave multivariable (longitudinal) design provided a potentially useful approach to the question of whether prior Silhouette ratings have any causal effect on subsequent self-concept beyond the substantial impact of prior self-concept. This approach has been used with considerable success in self-concept research to demonstrate that academic self-concept has a positive effect on subsequent academic achievement beyond the effect of prior academic achievement (Byrne, 1996a; Marsh, 1990a, 1993a; Marsh & Yeung, 1997). The critical prediction was that Time 1 (T1) Silhouette ratings contribute to Time 2 (T2) self-concept ratings beyond the contribution of T1 self-concept ratings. Results of these multiwave-multivariable causal models indicated that prior (T1) Actual and Ideal factors influenced subsequent (T2) self-concepts beyond the effects of prior self-concepts, arguing for the causal effects of Actual and Ideal body image on self-concept. Consistent with cognitive discrepancy models, these results imply that higher Actual selves and lower Ideal selves lead to higher subsequent self-concepts.

The chameleon effect

Self-esteem items are assumed to measure a unidimensional construct that does not depend upon the other items with which they appear. [e.g., Overall I have a lot to be proud of; Most things I do I do well; Overall I'm a failure; Nothing I do seems to turn out right]. In contrast, the chameleon effect posits Esteem responses assume the nature of the immediate context (e.g.,

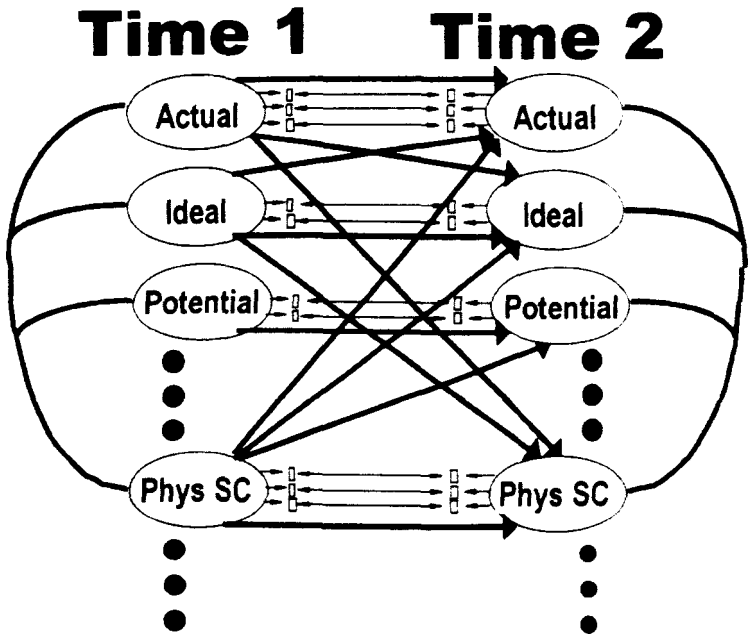

Figure 1

In this model only paths from Actual and Ideal silhouette factors to the self-concepts factors are posited (paths from Future and Potential silhouette ratings are constrained to be zero). Only 4 of 11 factors are presented in order to avoid clutter (all factors 11 factors – 4 silhouette factors and 7 self-concept factors – were considered at T1 and T2). The most important predictions are that the path from Actual-self to physical self-concept factors are **positive** whereas the path from Ideal-self to physical self-concept factors are **negative**. [Ovals represent latent factors. Boxes represent the indicators associated with each factor. Dark, single-headed arrows are path coefficients leading from one set of latent factors to another set of latent factors. Curved lines connecting the latent factors are correlations among factors. Light single-headed lines from each latent factor (the ovals) to its indicators (the boxes) are factor loadings. Light double-headed arrows between matching T1 and T2 indicators represent correlated uniquenesses associated with the same indicators administered at T1 and T2.]

Esteem items in a Physical Self-concept instrument will be more "physical" than the same Esteem items in an Academic Self-concept instrument). Three confirmatory factor analysis studies showed that the same Esteem items embedded in different instruments measured distinct factors. In a series of three studies, Marsh and Yeung (1999) demonstrated that the same Esteem items were:

- more "physical" in a Physical Self-concept instrument.
- more "academic" in an Academic Self-concept instrument.
- more "artistic" in a Performing Arts Self-concept instrument.

Results have theoretical and practical implications for use of Esteem in correlational and experimental studies. So called Esteem responses in sport/exercise contexts may really measure Global Physical Self-concept.

Elite athlete and nonelite groups

Is the PSDQ suitable for elite athletes? Marsh (1998) evaluated age and gender differences in responses to the PSDQ by elite athletes and nonathlete groups (N = 1,514).

Mean differences:

- PSDQ responses were much higher for two elite athletes groups (Australian Institute of Sport and Sports High School) than for two nonelite groups (nonathletes at Sports High School and Nonsport High School).
- Males had higher Physical Self-concepts than females. Gender differences, however, were smaller for elite athletes.
- Students in a nonsports high school had higher Physical Self-concepts than the nonelite athletes in an athletically selective high school. This is consistent with social comparison theory because the elite athletes provide a demanding frame of reference.

Factor Structure:

- CFA demonstrated the 11 PSDQ factors for all four groups. Factor loadings were invariant across the four groups.
- Factor variances and correlations were invariant across the two elite groups and the two nonelite groups.
- As predicted, PSDQ factors were more distinct (smaller *r*s) for elite athletes.
- Surprisingly, relations between Global Esteem and the PSDQ scales were no higher for elite athletes than the nonelite groups.
- The results demonstrate the appropriateness of the PSDQ for elite athletes, extend theoretical understanding of self-concept, and illustrate the power and flexibility of CFA.

A multicohort-multioccasion design to evaluation of gender and age effects in physical self-concept

The purpose of this study (Marsh, 1998) was to demonstrate the use of a multicohort-multioccasion design to evaluate gender and age effects in physical self-concept for elite-athletes and nonathletes. Elite athletes and nonathletes attending a highly selective sports high school completed the PSDQ on four occasions over a two-year period. Initially, the psychometric properties of PSDQ responses were evaluated as well as their generalizability across the two samples and over time. Based on previous research (e.g., Marsh, 1989; Marsh et al., 1995; Marsh & Craven, 1997), physical self-concepts were predicted to be substantially higher for elite athletes than nonathletes and substantially higher for men than women, but gender differences were predicted to be smaller for elite athletes than nonathletes. Research based on adolescent nonathletes responses to the SDQ physical ability and appearance scales suggests that there are small negative linear and U-shaped quadratic effects, but there is no basis for predicting whether a similar pattern of results will be evident for elite athletes.

Due in part to limitations in existing self-concept research, reviewers (e.g., Byrne, 1996a; Crain, 1996; Marsh, Craven & Debus, 1991, 1998; Wylie, 1979, 1989) emphasized the need for longitudinal studies more appropriate for evaluating the development of self-concept. In particular, studies of age and gender effects are typically based on a single wave of data from multiple age cohorts so that substantive and developmental implications relied primarily on cross-sectional comparisons. Here, in a multicohort-multioccasion design (Marsh, Craven, & Debus, 1998; also see Baltes & Nesselroade, 1979), four waves of physical self-concept responses were collected over a two-year period for the same high-school students from four age cohorts. Based on the multicohort-multioccasion design, it is possible to cross-validate cross-sectional (multiple age cohort) comparisons and true longitudinal (multiple occasion) comparisons. Longitudinal differences across all cohorts provide the longitudinal tests. Cohort differences across all waves reflect the cohort differences. The cohort x occasion interactions provide a test of whether one of these comparisons

depends on the other. Like any other validation procedure, if there is no reasonable agreement between different approaches (longitudinal and cross-sectional comparisons in this application), then there is need to explore further the substantive basis of the differences. Hence, this design provides a much stronger basis for evaluating age and gender developmental differences. In particular, if there are generalizable developmental differences in physical self-concept, then it is expected that differences associated with between-group (cross-sectional) age cohorts should be consistent with differences associated with within-group (longitudinal) comparisons over time. Hence, the present study extends previous self-concept research by focusing specifically on age and gender effects in multiple dimensions of physical self-concept and by demonstrating a potentially stronger methodological design for evaluating these issues that should have broad applicability in sport/exercise research.

Age-related differences due to gender, athletic group, year in school cohorts, and their interaction were largely consistent with predictions based on previous research. Across all 10 physical self-concepts, there were substantial differences due to group (athletes > nonathletes), gender (males > females), and gender x group interactions (gender differences smaller for athletes than nonathletes). There were no significant effects of age cohort (year in school) and only very small effects of occasions. Thus, longitudinal and cross-sectional comparisons agreed in showing that mean levels of physical self-concept were stable over this potentially volatile adolescent period and that this stability generalized over gender, age, and the athlete groups. These results also supported the construct validity of PSDQ interpretations based on the known-group difference approach (i.e., *a priori* predictions that particular known groups has higher or lower scores are supported). Even the relative sizes of the athlete group effects for the different PSDQ scales were reasonable with the largest effects (over 10% of the variance explained) for physical activity, endurance, sport competence, and coordination and the smallest effects (less than 2% of variance explained) for health, appearance, and body fat. The consistent athlete group x gender interactions are a potentially important contribution. Whereas elite athletes had systematically higher physical self-concepts than nonathletes, this advantage was larger for women than men in this adolescent period. The design and methodology of present investigation provided a strong basis for the evaluation of these findings. In particular, both the relatively smaller gender differences for elite-athletes and the relatively larger gender differences for nonathletes were relatively consistent across the cross-sectional comparisons over the four age cohorts and true longitudinal comparisons across the four testing occasions within each year.

A potentially important contribution of the present investigation was the demonstration of the multicohort-multioccasion design. In particular, the juxtaposition of the age effects based on the (cross-sectional) age cohort comparisons and the true longitudinal comparisons based on multiple occasions within each age cohort provided an important basis for cross-validating interpretations based on these two alternative measures of developmental effects. In the present investigation, for example, the very small cross-sectional effects of year in school (age cohorts) were reasonably consistent with the very small true longitudinal effects due to time (multiple occasions). It is, of course, important to emphasize that the multicohort-multioccasion design is not only relevant to physical self-concept development, but should have broad applicability for sport/exercise researchers who are interested, for example, in evaluating development of physical attributes, physical fitness, or physical skills.

An intervention to enhance multiple physical self-concepts, activity and health-related fitness

Let me now summarize results of some

recent research that shows a future direction for physical self-concept research. The enhancement of physical self-concept is a desirable goal but it is also postulated to facilitate long-term physical activity, physical fitness, and health-related behaviors. This research examines the mutually reinforcing benefits of simultaneously improving Physical Self-concept and health-related physical fitness.

Intervention: Design

The intervention was conducted with high school students randomly assigned to

- aerobics classes,
- circuit training classes,
- or a control group.

The major outcome measures were:

- physical fitness (standardized tests of fitness),
- physical activity (self-reported levels),
- multiple components of Physical Self-concept, and
- body composition (body mass, girths, and technical measures of body fat).

All students were tested:

- prior to the start of the intervention (pretest),
- shortly after the completion of the intervention (posttest), and
- 3 months after the end of the intervention (follow up).

Physical Self-concept/Fitness Intervention: Results

The major findings were that the two experimental groups experienced significant improvements in all the major outcome variables when compared to results from the control group. More specifically, on average, students participating in the intervention:

- were *more physically fit* based on tests of flexibility, endurance, strength and power;
- had *improved body composition* as assessed by body mass index, body girths and skinfolds;
- were *more physically active;*
- *improved Physical Self-concept* in multiple areas.
- Gains were similar for males and females and across the two intervention groups.
- The intervention was most beneficial for students who were initially least fit, most overweight, and least physically active.

Elite Athlete SDQ (EASDQ)

The PSDQ is suitable for elite athletes. There may, however, be other components of Physical Self-concept that are particularly relevant for elite athletes. We have recently developed the Elite Athlete Self-Description Questionnaire (EASDQ; see Marsh, Hey, Roche, & Perry, 1997; Marsh, Hey, Johnson, & Perry, 1997). The content of the EASDQ scales is summarized in Appendix 1. The EASDQ was administered to elite athletes from a selective sports high school (N = 349) and from the Australian Institute of Sport (N = 151).

- CFA of responses by the total group identified the six a *priori* factors .
- Multiple group CFAs of responses by each separate group (AIS and elite high school athletes) supported the factorial invariance of responses across the two groups.
- Hierarchical CFA provided good support for a single higher-order factor and the invariance of the hierarchical structure across two groups.

Results support the appropriateness of the EASDQ for diverse groups of elite athletes. More research is needed, however, relating EASDQ responses to external validity criteria like those used in PSDQ research and to criteria that are more specific to elite athletes (e.g., actual performance in competition).

Elite performers and highly selective environments

In research pursued in collaboration with the

Australian Institute of Sport and the Australian Federation of Swimming, I will measure elite athlete self-concepts of all participants in the upcoming PANPAC international swimming meeting (including competitors from more than 30 countries). Participants will complete instruments prior to and following the competition. Also available are world rankings and previous "personal bests" of each athlete. Of particular interest will be how physical self-concepts vary as a function of performance (absolute and relative) and country, and how this influences subsequent performance.

Summary and Implications

Interest in physical self-concept stems from its recognition as a valued outcome, its role as a moderator variable, interest in its relation with other constructs, and concerns with methodological and measurement issues. Theory, measurement, research and practice are inexorably intertwined; each will suffer if one is ignored. Research described represents an interplay between theory and empirical research. It supports the construct validity approach that guided PSDQ research. The strongest contribution of PSDQ research may be the development of instruments, based on strong empirical and theoretical foundations, for the measurement of multiple dimensions of physical self-concept. The research also outlines a construct validity approach in which an emphasis on good measurement is a critical feature of good research. This approach should be useful to other areas of sport psychology and to sports sciences more generally.

References

Baltes, P. B., & Nesselroade, J. R. (Eds.). (1979). *Longitudinal research in the study of behavior and development*. New York: Academic.

Boyle, G. J. (1994). Self-Description Questionnaire II: A review. *Test Critiques, 10,* 632-643.

Bracken, B. A. (1992). *Multidimensional Self-Concept Scale*. Austin, TX: Pro-ed.

Burns, R. B. (1979). *The self-concept: Theory, measurement, development and behaviour*. London: Longman.

Byrne, B. M. (1984). The general/academic self-concept nomological network: A review of construct validation research. *Review of Educational Research, 54,* 427-456.

Byrne, B. M. (1996a). Academic self-concept: Its structure, measurement, and relation to academic achievement. In B. A. Bracken (Ed.), *Handbook of self-concept* (pp. 287-316). New York: Wiley.

Byrne, B. (1996b). *Measuring self-concept across the life span: Issues and instrumentation*. Washington, DC: American Psychological Association.

Coopersmith, S. A. (1967). *The antecedents of self-esteem*. San Francisco: Freeman.

Crain, R. M. (1996). The influence of age, race, and gender on child and adolescent multidimensional self-concept. In B. A. Bracken (Ed.), *Handbook of self-concept: Developmental, social, and clinical considerations* (pp. 395-420). New York: Wiley.

Ellis, L., Marsh, H. W., & Craven, R. G. (2000, October). *The self-concept of preschool children: Measurement and multidimensional structure*. Paper presented at the 2000 Self-concept Enhancement and Learning Facilitation (SELF) Conference, Sydney, Australia.

Fitts, W. H. (1964). *Tennessee Self-Concept Scale: Test Booklet*. Nashville, TN: Counsellor Recordings and Tests.

Fleishman, F. A. (1964). *The structure and measurement physical fitness*. Englewood Cliffs, NJ: Prentice-Hall.

Fleming, J. S., & Courtney, B. E. (1984). The dimensionality of self-esteem: II. Hierarchical facet model for revised measurement scales.

Journal of Personality and Social Psychology, 46, 404-421.

Fox, K. R. (1990). *The Physical Self-Perception Profile Manual.* DeKalb, IL: Office for Health Promotion, Northern Illinois University.

Fox, K. R., & Corbin, C. B. (1989). The Physical Self-Perception Profile: Development and preliminary validation. *Journal of Sport & Exercise Psychology, 11*, 408-430.

Gill, D. L., Dzewaltowski, D. A., & Deeter, T. E. (1988). The relationship of competitiveness and achievement orientation to participation in sport and nonsport activities. *Journal of Sport & Exercise Psychology, 10*, 139-150.

Harter, S. (1985). *Manual for the Self-perception Profile for Children*. Denver, CO: University of Denver.

Harter, S., & Pike, R. (1984). The pictorial scale of perceived competence and social acceptance for young children. *Child Development, 55*, 1969-1982.

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

James, W. (1890/1963). *The principles of psychology*. New York: Holt, Rinehart & Winston.

Keith, L. K., & Bracken, B. A. (1996). Self-concept instrumentation: A historical and evaluating review. In B. A. Bracken (Ed.), *Handbook of self-concept: Development, social and clinical considerations* (pp. 91-170). New York: Wiley.

Marsh, H. W. (1988). Multitrait-multimethod analyses. In J. P. Keeves (Ed.), *Educational research methodology, measurement and evaluation: An international handbook* (pp. 570-580). Oxford: Pergamon Press.

Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to adulthood. *Journal of Educational Psychology, 81*, 417-430.

Marsh, H. W. (1990a). Confirmatory factor analysis of multitrait-multimethod data: The construct validation of multidimensional self-concept responses. *Journal of Personality, 58*, 661-692.

Marsh, H. W. (1990b). A multidimensional, hierarchical self-concept: Theoretical and empirical justification. *Educational Psychology Review, 2*, 77-171.

Marsh, H. W. (1990c). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology, 82*, 623-636.

Marsh, H. W. (1993a). Academic self-concept: Theory measurement and research. In J. Suls (Ed.), *Psychological perspectives on the self* (Vol. 4, pp. 59-98). Hillsdale, NJ: Erlbaum.

Marsh, H. W. (1993b). The multidimensional structure of physical fitness: Invariance over gender and age. *Research Quarterly for Exercise and Sport, 64*, 256-273.

Marsh, H. W. (1993c). Physical fitness self-concept: Relations to field and technical indicators of physical fitness for boys and girls aged 9-15. *Journal of Sports and Exercise Psychology, 15*, 184-206.

Marsh, H. W. (1994). Sport motivation orientations: Beware of the jingle-jangle fallacies. *Journal of Sport and Exercise Psychology, 16*, 365-380.

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.

Marsh, H. W. (1996b). Physical Self-Description Questionnaire: Stability and discriminant validity. *Research Quarterly for Exercise and Sport, 67*, 249-264.

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 self-concepts for adolescent elite-athletes and non-athletes: A multi-cohort-multi-occasion design. *Journal of Sport & Exercise Psychology, 20*(3), 237-259.

Marsh, H. W. (1999). Cognitive discrepancy

models: Actual, ideal, potential, and future self-concept perspectives of body image. *Social Cognition, 17*, 46-75.

Marsh, H. W., Asci, F. H., & Marco, I. T. (2002). Multitrait-multimethod analyses of two physical self-concept instruments: A cross-cultural perspective. *Journal of Sports and Exercise Psychology, 24*, 99-119.

Marsh, H. W., Byrne, B. M., & Shavelson, R. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology, 80*, 366-380.

Marsh, H. W., & Craven, R. (1997). Academic self-concept: Beyond the dustbowl. In G. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement, and adjustment* (pp. 131-198). Orlando, FL: Academic.

Marsh, H. W., Craven, R. 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.

Marsh, H. W., Craven, R. G., & Debus, R. (1998). Structure, stability, and development of young children's self-concepts: A multi-cohort-multioccasion study. *Child Development, 69*(4), 1030-1053.

Marsh, H. W., & Gouvernet, P. (1989). Multidimensional self-concepts and perceptions of control. Construct validation of responses by children. *Journal of Educational Psychology, 81*, 57-69.

Marsh, H. W., & Hattie, J. (1996). Theoretical perspectives on the structure of self-concept. In B. A. Bracken (Ed.), *Handbook of self-concept* (pp. 38-90). New York: Wiley.

Marsh, H. W., Hey, J., Johnson, S., & Perry, C. (1997). Elite Athlete Self-description Questionnaire: Hierarchical confirmatory factor analysis of responses by two distinct groups of elite athletes. *International Journal of Sport Psychology, 28*, 237-258.

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.

Marsh, H. W., & Jackson, S. A. (1986). Multidimensional self-concepts, masculinity and femininity as a function of women's involvement in athletics. *Sex Roles, 15*, 391-416.

Marsh, H. W., Marco, I. T., & Asci, F. H. (in press). Cross-cultural Validity of the Physical Self Descriptin Questionnaire: Comparison of factor structures in Australia, Spain and Turkey. *Research Quarterly for Exercise and Sport.*

Marsh, H. W., & McDonald-Holmes, I. W. (1990). Multidimensional self-concepts: Construct validation of responses by children. *American Education Research Journal, 27*, 89-117.

Marsh, H. W., & Peart, N. (1988). Competitive and cooperative physical fitness training programs for girls: Effects on physical fitness and on multidimensional self-concepts. *Journal of Sport and Exercise Psychology, 10*, 390-407.

Marsh, H. W., Perry, C., Horsely, C., & Roche, L. A. (1995). Multidimensional self-concepts of elite athletes: How do they differ from the general population? *Sport and Exercise Psychology, 17*, 70-83.

Marsh, H. W., & Redmayne, R. S. (1994). A multidimensional physical self-concept and its relation to multiple components of physical fitness. *Journal of Sports and Exercise Psychology, 16*, 45-55.

Marsh, H. W., & Richards, G. (1988a). The Outward Bound Bridging Course for low achieving high-school males: Effect on academic achievement and multidimensional self-concepts. *Australian Journal of Psychology, 40*, 281-298.

Marsh, H. W., & Richards, G. E. (1988b). The Tennessee Self-Concept Scales: Reliability, internal structure, and construct validity. *Journal of Personality and Social Psychology, 55*, 612-624.

Marsh, H. W., Richards, G., & Barnes, J. (1986a). Multidimensional self-concepts: A long term follow-up of the effect of participation in an Outward Bound program. *Personality and Social Psychology Bulletin, 12*, 475-492.

Marsh, H. W., Richards, G., & Barnes, J. (1986b). Multidimensional self-concepts: The effect of participation in an Outward Bound program. *Journal of Personality and Social Psychology, 45*, 173-187.

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. *Sport and Exercise Psychology, 16*, 270-305.

Marsh, H. W., & Roche, L. A. (1996). Predicting self-esteem from perceptions of actual and ideal ratings of body fatness – is there only one ideal supermodel? *Research Quarterly for Exercise & Sport, 67*(1), 13-23.

Marsh, H. W., & Shavelson, R. J. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist, 20*, 107-125.

Marsh, H. W., & Smith, I. D. (1982). Multitraitmultimethod analyses of two self-concept instruments. *Journal of Educational Psychology, 74*, 430-440.

Marsh, H. W., & Yeung, A. S. (1997). Coursework selection: The effects of academic self-concept and achievement. *American Educational Research Journal, 34*, 691-720.

Marsh, H. W., & Yeung, A. S. (1999). The liability of psychological ratings: The chameleon effect in global self-esteem. *Personality and Social Psychology Bulletin, 25*, 49-64.

Marx, R. W., & Winne, P. H. (1978). Construct interpretations of three self-concept inventories. *American Educational Research Journal, 15*, 99-108.

Nelson, J. K. (1989). Measurement methodology for affective tests. in M. J. Safrit & T. M. Wood (Eds.), *Measurement concepts in physical education and exercise science* (pp. 271-295). Campaign, IL: Human Kinetics Books.

Ostrow, A. C. (1990). *Directory of psychological tests in the sport and exercise sciences*. Morgantown, WV: Fitness Information Technology.

Richards, G. E. (1987, January). *Outdoor education in Australia in relation to the Norman Conquest, a Greek Olive Grove and the External Perspective of a Horse's Mouth*. Paper presented at the 5th National Outdoor Education Conference, Perth, Western Australia.

Richards, G. E. (1988). *Physical Self-Concept Scale*. Sydney: Australian Outward Bound Foundation

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Validation of construct interpretations. *Review of Educational Research, 46*, 407-441.

Sonstroem, R. J. (1978). Physical estimation and attraction scales: Rationale and research. *Medicine and Science in Sport, 10*, 97-102.

Sonstroem, R. J. (1988). *Psychological models in exercise adherence: Its impact on public health*. Champaign, IL: Human Kinetics Books.

Vealey, R. S. (1986). Conceptualization of sport-confidence and competitive orientation: Preliminary investigation and instrument development. *Journal of Sport Psychology, 8*, 221-246.

Wells, L. E., & Marwell, G. (1976). *Self-esteem: Its conceptualization and measurement*. Beverly Hills, CA: Sage.

Wylie, R. C. (1974). *The self-concept* (Rev. ed., Vol. 1). Lincoln, NE: University of Nebraska Press.

Wylie, R. C. (1979). *The self-concept* (Vol. 2). Lincoln, NE: University of Nebraska Press.

Wylie, R. C. (1989). *Measures of self-concept*. Lincoln, NE: University of Nebraska Press.

Appendix 1
Elite Athlete Self-Description Questionnaire (EADQ):
Summary of the content of the six scales

Scale	Summary
Skills:	In my best sport/event I am a skillful athlete: My technical skills are better than most at my level of competition; I excel because of my skill level.
Body:	I excel in my best sport/event because of the suitability of my body composition, body shape, body structure; Having the right body helps me perform well.
Physiological Competence (aerobic):	In my best sport/event I am aerobically superior compared to my team mates/competitors; My capacity for endurance makes me a good performer; Coaches and my competitors see me as very fit aerobically.
Physiological Competence (anaerobic):	In my best sport/event I am anaerobically superior compared to others; My capacity for short bursts of high intensity activity makes me a good performer; Coaches and my competitors see my as very fit anaerobically.
Mental Competence:	I have better mental skills, commitment, discipline, focus, emotional control than others at my level in my best sport/event.
Overall Performance:	Excellent performer; perform to my ability level; give peak performance when necessary; can "pull it all together".