SELF DESCRIPTION QUESTIONNAIRE - II

SDQ II

MANUAL

HERBERT W. MARSH

SELF DESCRIPTION QUESTIONNAIRE - II

SDQ II

MANUAL

HERBERT W. MARSH

Acknowledgments

I would like to thank the many colleagues and students who have contributed to the research described in this Manual. In particular I would like to acknowledge the support of Ian Smith, Jennifer Barnes, Raymond Debus, Samuel Ball, John Parker and Barbara Byrne. Above all, I would like to thank Richard Shavelson, whose theoretical model of self-concept was the basis for the instrument and research described in this Manual and who provided support and critical evaluation of the research.

Herbert W. Marsh
August 1990

Copyright:

Herbert W. Marsh
University of Western Sydney, Macarthur
P. O. Box 555, Campbelltown N.S.W. 2560, Australia

Contents

Tables vii
Figures ix

Chapter 1. Introduction 1
Theoretical Basis of the SDQ 1
General Description 1
Overview of Empirical Research 6
Reliability and Factor Structure 6
Construct Validity 7
Changes in Self-Concept 7

Chapter 2. Administration and Scoring 9
Administration Instructions 9
Standard Administration Procedure 9
Read-Aloud Administration Procedure 10
Scoring Instructions 11
Calculating Individual Raw Scores 11
Calculating a Total Raw Score 13
Converting Raw Scores to Scaled Scores 14
Factor Scores 17

Chapter 3. Norms Development and Interpretation 19
Scale Development 19
Standardization 19
Sample Description 19
Derivation of Norms 20
Interpretation 20
Individual Scales 21
General Self Scale 23
Total Self-Concept Score 23

Chapter 4. Reliability 25
Internal Consistency 25
Standard Error of Measurement 26
Stability of SDQ-II Responses Over Time 26

Chapter 5. Construct Validity 31
The Shavelson et al. Model 32
Factor Analytic Studies 33
Use of Item Pairs 33
Results of the Total Group Factor Analysis 34
Similarity of the Factor Structures Across Sex and Age 37
Hierarchical Structure of Self-Concept 38
Summary and Implications 39

Chapter 6. Age and Sex Effects 41
Review of the Literature 41
Age Effects in Self-Concept 41
Sex Effects in Self-Concept 44
Age Effects in the Structure and Hierarchy of Self-Concept 46
Sex and Age Effects in SDQ-I, SDQ-II, and SDQ-III Responses 48
Main and Interaction Effects 48
Age Effects on Differentiation Among Self-Concept Facets 53
Summary and Implications 55

Chapter 7. Self-Concept and Academic Achievement 59
Relationship of Self-Concept to Academic and Other Constructs 59
Educational Coping Skills 59
Math Achievement and Physical Fitness 62
Measures of Academic Achievement 63
Internal–External (I–E) Frame of Reference Model 66
Relationship of Verbal and Math Self-Concepts 66
Empirical Support for the I–E Model 68
Generality of the I–E Model 68
Summary and Implications 74

Chapter 8. Effects of Interventions 75
Study of the Outward Bound Standard Course 76
Study of the Outward Bound Bridging Course 76
Effects of Competitive and Cooperative Physical Fitness Training Programs 78
Summary and Implications 85

Chapter 9. Relationship of Self-Concept to Masculinity, Femininity, and Androgyny 87
The Androgyny Construct 87
Theoretical Models 88
Empirical Support for the Models 89
Relation Between Masculine/Feminine Identity and SDQ-II 89
Summary and Implications 93

Chapter 10. Summary and Overview 95

Appendices 97
References 111

Tables

Table 1. Scale Descriptions 2

Table 2. Internal Consistency and Standard Errors of Measurement (SE_Ms) 25

Table 3. Item Statistics by Scale 27

Table 4. Internal Consistency and Test–Retest Reliability over a 7-Week Interval 30

Table 5. Factor Analysis for the SDQ-II Normative Sample 35

Table 6. Factor Analyses of SDQ-II Responses for the Total Group and Four Subgroups 38

Table 7. Sex and Age Effects (%Variance Explained) 47

Table 8. Summary of Sex and Age Effects (%Variance Explained) for SDQ-I, -II, and -III on Raw Scale Scores 50

Table 9. Differentiation of Self-Concept with Age: SDQ-I, -II, and -III 54

Table 10. Correlations Between the SDQ-II Scales and Selected Academically Related Variables and Anxiety 61

Table 11. Correlations of SDQ-II Scales with Measures of Math Achievement and Physical Fitness 62

Table 12. Correlations Among Self-Concept Scores, Academic Achievement, and IQ 64

Table 13. Standardized Beta Weights from the Multiple Regressions Relating Academic Self-Concepts to Academic Achievement 65

Table 14. Correlations Among Variables Considered in Tests of the I–E Model 70

Table 15. Effects of Participation in Alternative Physical Fitness Training Programs on SDQ-II Responses and on Physical Fitness 82

Table 16. Intervention Effects: Discriminant Analysis of Multivariate Group Differences on the Set of Outcome Measures Using Backward Elimination for Nonsignificant Variables 84

Table 17. Correlations and Beta Weights for Four M/F Scores and SDQ-II Responses 91

Table 18. Correlations Among Bipolar M/F Scores, Gender, and SDQ-II Responses 93

Appendix A-1. Percentiles and *T* Scores for the Total Normative Sample 98

Appendix A-2. Percentiles and *T* Scores: Males 100

Appendix A-3. Percentiles and *T* Scores: Females 102

Appendix A-4. Percentile Equivalents of Total Self-Concept Raw Scores for the Total, Male, and Female Samples 104

Appendix A-5. *T*-Score Equivalents of Total Self-Concept Raw Scores for the Total, Male, and Female Samples 106

Appendix B. Factor Score Coefficients (FSC) Based on the Total Normative Sample 108

Figures

Figure 1. Physical Abilities Scale 3

Figure 2. Physical Appearance Scale 3

Figure 3. Opposite-Sex Relations Scale 3

Figure 4. Same-Sex Relations Scale 4

Figure 5. Parent Relations Scale 4

Figure 6. Honesty–Trustworthiness Scale 4

Figure 7. Emotional Stability Scale 5

Figure 8. Math Scale 5

Figure 9. Verbal Scale 5

Figure 10. General School Scale 6

Figure 11. General Self Scale 6

Figure 12. Sample Page from the SDQ-II Questionnaire 12

Figure 13A. Calculating Individual and Total Self-Concept Raw Scores 15

Figure 13B. Plotting a *T*-Score Profile and Percentile–Rank Equivalents 16

Figure 14. Age and Sex Effects on the Scales Common to the Three SDQ Instruments 51

Figure 15. Path Model of Relationships Among Achievements and Reading and Math Self-Concepts: Internal–External Frame of Reference Model 67

Figure 16A. Basic Internal–External Frame of Reference Model 72

Figure 16B. Internal–External Frame of Reference Model with Gender 72

Figure 16C. Internal–External Frame of Reference Model with School Achievement 73

Figure 16D. Internal–External Frame of Reference Model with Gender and School Achievement 73

Chapter 1. Introduction

Theoretical Basis of the SDQ

Self-concept has been used to explain a wide range of behaviors. Moreover, the goal of fostering positive self-concept has been the focus of a variety of educational and clinical interventions. Despite its importance, reviews of the literature on self-concept have noted important shortcomings, especially the lack of a sound theoretical basis and the poor quality of instrumentation used to measure it (Burns, 1979; Shavelson, Hubner, & Stanton, 1976; Wells & Marwell, 1976; Wylie, 1974, 1979). To remedy this situation, Shavelson et al. proposed a multifaceted, hierarchical model of self-concept (see Chapter 5). This model served as the basis for the *Self-Description Questionnaire-II* (SDQ-II) and its two companion instruments, the SDQ-I (March, 1988) and SDQ-III (March, in press).

The SDQ-II is designed to measure self-concept in younger adolescents (i.e., junior and senior high school students in grades 7 through 10). The original instrument, SDQ-I, used to measure self-concept in elementary-aged children in grades 4 through 6, is specifically designed to measure three areas of academic self-concept (Reading, Mathematics, and General School) and four areas of nonacademic self-concept (Physical Abilities, Physical Appearance, Peer Relations, and Parent Relations), all derived from the Shavelson et al. model. Subsequently, the SDQ-I was revised to include a General Self scale (see the SDQ-I Manual). The SDQ-II contains the seven scales from the original SDQ-I, although the Peer scale is divided into Same-Sex and Opposite-Sex Relations scales. In addition, it includes two new scales—Emotional Stability and Honesty-Trustworthiness—and a General Self scale.

The identification of theoretically consistent and distinguishable facets of self-concept is a prerequisite to the study of how self-concept relates to other constructs. Consequently, early research with the SDQ focused on the internal characteristics of self-concept, particularly its facets and their organization. More recent SDQ research has focused on the relationships between specific facets of self-reported self-concept and other constructs (such as academic achievement and inferred self-concepts based on ratings by significant others) and on the effects of interventions designed to alter self-concept.

General Description

The 102-item SDQ-II (see Table 1) assesses three areas of academic self-concept, seven areas of nonacademic self-concept, and general self-concept derived from the Rosenberg (1965, 1979) self-esteem scale. These 11 scales, which are also summed to yield a Total Self-Concept score, reflect an adolescent's self-ratings in various areas of self-concept. (For a list of the items composing each scale, see Figures 1 through 11.)

In completing the SDQ-II, adolescents are asked to respond to simple declarative sentences (e.g., "I am good looking," "I worry a lot," "I have trouble with most academic subjects") with one of six responses: False; Mostly False; More False Than

Table 1. Scale Descriptions

Scale	Description	Number of Items
Physical Abilities	Skills and interest in sports and physical activities	8
Physical Appearance	Physical attractiveness	8
Opposite-Sex Relations	Interactions with peers of the opposite sex	8
Same-Sex Relations	Interactions with peers of the same sex	10
Parent Relations	Interactions with parents	8
Honesty–Trustworthiness	Truthfulness and dependability	10
Emotional Stability	Emotional well being and freedom from psychopathology	10
Math	Ability, enjoyment, and interest in mathematics and reasoning	10
Verbal	Ability, enjoyment, and interest in English and reading	10
General School	Ability, enjoyment, and interest in school subjects	10
General Self	Self-worth, self-confidence, self-satisfaction	10

True; More True Than False; Mostly True; or True. Each of the 11 SDQ-II facet scores is based on the adolescent's self-ratings on 8–10 items, half of which are negatively worded in order to disrupt positive response biases. In Figures 1–11, the positive items are indicated with a plus sign (+), the negative with a minus sign (−).

The SDQ-II can be administered either individually or in groups. Although appropriate interpretation requires some familiarity with basic psychometric principles, no special training is required to administer the questionnaire. The SDQ-II Questionnaire is designed to be self-explanatory and may be completed by respondents without any additional instructions (e.g., as part of a mail survey). Most adolescents complete the SDQ-II in less than 20 minutes, although some respondents may take a little longer. For group administration, the administrator reads aloud the instructions on the front of the instrument before each adolescent completes the instrument. (See Chapter 2 for complete administration instructions.)

The SDQ-II is intended for use by junior high and high school students in grades 7 through 12. Although the SDQ-II was specifically designed for these grades, some studies have used the SDQ-III with students in grades 9–10 (Marsh & O'Niell, 1984) or the SDQ-I with students in grades 7–10 whose verbal skills are below average (Marsh & Richards, 1988a). In addition, the SDQ-III can—and has—been used with students in their last two years of high school, but the SDQ-III is designed primarily for college-aged students. The SDQ-II is also appropriate for older adolescents and young adults, although important components of adult life are not included.

Norms based on the responses of 5,494 Australian students are presented for each of the SDQ-II scales and for the Total Self-Concept score. Norms are presented as both percentiles and nonnormalized *T* scores (see Appendix A). Because SDQ-II responses vary according to sex (see Chapter 6), separate norms tables are presented for males and females.

Item No.	Item
5.	I enjoy things like sports, gym and dance. (+)
16.	I am lazy when it comes to things like sports and hard physical exercise. (−)
27.	I'm good at things like sports, gym and dance. (+)
38.	I am awkward at things like sports, gym and dance. (−)
49.	I'm better than most of my friends at things like sports, gym and dance. (+)
60.	I try to get out of sports and physical education classes whenever I can. (−)
71.	I can run a long way without stopping. (+)
82.	I hate things like sports, gym and dance. (−)

Figure 1. Physical Abilities Scale

Item No.	Item
2.	Nobody thinks that I'm good looking. (−)
13.	I have a nice looking face. (+)
24.	Most of my friends are better looking than I am. (−)
35.	I am good looking. (+)
46.	I hate the way I look. (−)
57.	Other people think I am good looking. (+)
68.	I am ugly. (−)
79.	I have a good looking body. (+)

Figure 2. Physical Appearance Scale

Item No.	Item
11.	People of the opposite sex whom I like don't like me. (−)
21.	I make friends easily with boys.**(+)
22.	I make friends easily with girls.*(+)
33.	I'm not very popular with members of the opposite sex. (−)
43.	I am popular with boys.**(+)
44.	I am popular with girls.*(+)
54.	I do not get along very well with boys.**(−)
55.	I do not get along very well with girls.*(−)
66.	I have lots of friends of the opposite sex.(+)
76.	Most boys try to avoid me.**(−)
77.	Most girls try to avoid me.*(−)
88.	I get a lot of attention from members of the opposite sex.(+)

* Items are scored only for males.
** Items are scored only for females.

Figure 3. Opposite-Sex Relations Scale

Item No.	Item
10.	It is difficult to make friends with members of my own sex.(–)
21.	I make friends easily with boys.*(+)
22.	I make friends easily with girls.**(+)
32.	Not many people of my own sex like me.(–)
43.	I am popular with boys.*(+)
44.	I am popular with girls.**(+)
54.	I do not get along very well with boys.*(–)
55.	I do not get along very well with girls.**(–)
65.	I have good friends who are members of my own sex.(+)
76.	Most boys try to avoid me.*(–)
77.	Most girls try to avoid me.**(–)
87.	I make friends easily with members of my own sex.(+)
95.	I have few friends of the same sex as myself.(–)
102.	I enjoy spending time with my friends of the same sex.(+)

* Items are scored only for males.
** Items are scored only for females.

Figure 4. Same-Sex Relations Scale

Item No.	Item
8.	My parents are usually unhappy or disappointed with what I do.(–)
19.	I get along well with my parents.(+)
30.	It is difficult for me to talk to my parents.(–)
41.	My parents treat me fairly.(+)
52.	I have lots of arguments with my parents.(–)
63.	My parents understand me.(+)
74.	I do not like my parents very much.(–)
85.	My parents really love me a lot.(+)

Figure 5. Parent Relations Scale

Item No.	Item
4.	I sometimes take things that belong to other people.(–)
15.	I am honest.(+)
26.	I sometimes tell lies to stay out of trouble.(–)
37.	I always tell the truth.(+)
48.	Cheating on a test is OK if I do not get caught.(–)
59.	Honesty is very important to me.(+)
70.	I sometimes cheat.(–)
81.	When I make a promise I keep it.(+)
91.	I often tell lies.(–)
98.	People can really count on me to do the right thing.(+)

Figure 6. Honesty–Trustworthiness Scale

Item No.	Item
7.	I am usually relaxed.(+)
18.	I worry more than I need to.(–)
29.	I don't get upset very easily.(+)
40.	I am often depressed and down in the dumps.(–)
51.	Other people get more upset about things than I do.(+)
62.	I am a nervous person.(–)
73.	I often feel confused and mixed up.(–)
84.	I get upset easily.(–)
93.	I am a calm person.(+)
100.	I worry about a lot of things.(–)

Figure 7. Emotional Stability Scale

Item No.	Item
1.	Mathematics is one of my best subjects.(+)
12.	I often need help in mathematics.(–)
23.	I look forward to mathematics classes.(+)
34.	I have trouble understanding anything with mathematics in it.(–)
45.	I enjoy studying for mathematics.(–)
56.	I do badly in tests of mathematics.(–)
67.	I get good marks in mathematics.(+)
78.	I never want to take another mathematics course.(–)
89.	I have always done well in mathematics.(+)
96.	I hate mathematics.(–)

Figure 8. Math Scale

Item No.	Item
6.	I'm hopeless in English classes.(–)
17.	I look forward to English classes.(+)
28.	I do badly on tests that need a lot of reading ability.(–)
39.	Work in English classes is easy for me.(+)
50.	I'm not very good at reading.(–)
61.	English is one of my best subjects.(+)
72.	I hate reading.(–)
83.	I get good marks in English.(+)
92.	I have trouble expressing myself when I try to write something.(–)
99.	I learn things quickly in English classes.(+)

Figure 9. Verbal Scale

Item No.	Item
9.	People come to me for help in most school subjects.(+)
20.	I'm too stupid at school to get into a good university.(–)
31.	If I work really hard I could be one of the best students in my school year.(+)
42.	I get bad marks in most school subjects.(–)
53.	I learn things quickly in most school subjects.(+)
64.	I am stupid at most school subjects.(–)
75.	I do well in tests in most school subjects.(+)
86.	I have trouble with most school subjects.(–)
94.	I'm good at most school subjects.(+)
101.	Most school subjects are just too hard for me.(–)

Figure 10. General School Scale

Item No.	Item
3.	Overall, I have a lot to be proud of.(+)
14.	Overall, I am no good.(–)
25.	Most things I do, I do well.(+)
36.	Nothing I do ever seems to turn out right.(–)
47.	Overall, most things I do turn out well.(+)
58.	I don't have much to be proud of.(–)
69.	I can do things as well as most people.(+)
80.	I feel that my life is not very useful.(–)
90.	If I really try I can do almost anything I want to do.(+)
97.	Overall, I'm a failure.(–)

Figure 11. General Self Scale

Overview of Empirical Research

The SDQ-II is one in a series of three instruments designed to measure self-concept in preadolescents (SDQ-I), young adolescents (SDQ-II), and late adolescents and young adults (SDQ-III). The initial instrument, the SDQ-I, and its extensive research basis are described in the *SDQ-I: Manual and Research Monograph* (Marsh, 1988). Because the SDQ-II and SDQ-III are based on the SDQ-I and its research, much of the background for these instruments is contained in this previous manual.

Reliability and Factor Structure

Within-network research, the identification of the salient components of self-concept and how they relate to each other, was the initial focus of all SDQ research including the initial studies with the SDQ-II. The main empirical tools for this type of research are reliability studies, factor analyses, and multitrait-multimethod designs. The coefficient alpha estimate of reliability (see Chapter 4) for each SDQ-II

scale is generally high, varying from .83 to .91 (median = .87), whereas the average correlation among the factors is modest (mean r = .18). Numerous factor analyses of SDQ-II responses have identified the factors that the instrument is designed to measure (see Chapter 5). Taken together, the studies reported in Chapters 4 and 5 demonstrate that the SDQ-II reliably measures distinct facets of self-concept, underscoring its usefulness as both a clinical and research instrument.

Construct Validity

As discussed previously, self-concept is a theoretical construct. Thus, it is appropriate to use a construct validation approach to test the validity of responses to the SDQ-II. Such an approach requires that SDQ-II responses be related to a wide variety of external criteria, and that each of its factors be significantly correlated with other constructs to which it is logically related and less correlated with other constructs to which it is logically unrelated. In validity research described in subsequent chapters, responses to the SDQ-II are related to several other variables including sex, age, academic achievement in particular subject areas, experimental interventions designed to enhance self-concept, and responses to other self-concept measures. SDQ-II responses are systematically related to these external criteria in a way that is consistent with the theory, thus supporting the construct validity of SDQ-II. Because the SDQ-II was derived from earlier research with the SDQ-I and SDQ-III and contains items from each of these instruments, support for the validity of SDQ-II is also inferred from the extensive research with the SDQ-I and SDQ-III. (See the SDQ-I Manual [Marsh, 1988] and the SDQ-III Manual [Marsh, in press].)

Changes in Self-Concept

Self-concept researchers and practitioners face an important dilemma in assessing changes in self-concept. From the perspective of measurement theory and, perhaps, mental health, it is important that self-concept be relatively stable over time. However, much of the interest in self-concept stems from purported changes in self-concept as a result of naturally occurring developmental or environmental phenomena, and particularly as a result of interventions specifically designed to enhance self-concept.

In fact, recent research suggests that self-concept *is* stable over time. Despite claims to the contrary, few interventions are sufficiently powerful to substantially affect self-concept. There are several reasons for these failures. Some interventions fail to show a measurable effect because they are inherently weak, others because they are applied to such a small number of subjects that changes cannot be reliably inferred, and still others because the self-concept instruments used do not measure specific components of self-concept that are logically related to the intended effects of the intervention. An important area of research with all three SDQ instruments discussed in this Manual is the study of the effects of interventions on responses to the SDQ and the examination of alternative explanations for these effects (see Chapter 8).

This Manual provides a comprehensive overview of the SDQ-II including its applications, development, and empirical research. Basic administration and scoring procedures are presented in Chapter 2. Chapter 3 describes the development and standardization of the instrument, followed by a discussion of different levels of interpretation. Chapters 4–6 probe the critical issues of reliability, construct validity, and the effects of moderator variables, especially age and sex. The relationships of general and more specific facets of self-concept to academic achievement and treatment interventions are discussed in Chapters 7 and 8.

Chapter 9 considers the relationship between reported self-concept and gender roles including masculinity, femininity, and androgyny. Finally, Chapter 10 presents a summary and overview of these results and provides a context for future research.

Chapter 2. Administration and Scoring

This chapter provides specific procedures for administering and scoring the SDQ-II. Administration directions are followed by instructions for calculating individual scale and total raw scores and for converting these scores into standard score metrics—percentiles and nonnormalized *T* scores.

Administration Instructions

The following administration procedures were used for the collection of SDQ-II data for the norm tables and research described in subsequent chapters. Hence, it is important to follow these procedures carefully to ensure that results are comparable to those described in the Manual. Although described for group administration, the procedures are essentially the same for individual administration.

Standard Administration Procedure

Instructions for completing the SDQ-II are printed on the first page of the Questionnaire and are readily understood by students in the age range for which the SDQ-II was designed.

- Give a copy of the SDQ-II Questionnaire (see Figure 12 for a sample page) and make sure that each student has a pencil with an eraser. (The SDQ-II may be completed in ink, but doing so makes it more difficult for students to change responses.) Ask the students to complete the identifying and background information at the top of the front page. *Make sure that none of the students opens the Questionnaire until instructed to do so.*

- Tell the students that their responses will be kept confidential and will not be made public. It is the responsibility of the examiner to honor this promise. If some aspect of this assurance is not applicable, it should be omitted, but these special circumstances should be noted. These circumstances may affect student responses.

- The instructions on the front of the Questionnaire should be read aloud before the students begin responding to the items. Ask the students to listen and follow along while you read aloud the instructions on the front page. Do not allow questions until after you have read the first sample item. Students are often confused at the end of the second paragraph, but this confusion usually clears up after the examples are given and explained. It may be useful to hold up the instrument when reading the third paragraph and to point to the six response blanks and headings before reading the material in parentheses after Example 1. Briefly pause after reading the instructions for Example 3 to allow students to mark their answers. Very few students have problems responding to Example 3, and most understand how to mark their answers. Nevertheless, make sure the

students understand how to respond. Before allowing students to begin, be sure that all questions have been answered.

- When you are ready to begin, say: **Turn over the page and begin. Once you have started, PLEASE DO NOT TALK.** Be sure to stop any talking, commenting, and deliberate or unconscious vocalization.

- If there are any questions, go to the individual student. If a student has trouble understanding a few words or expressions, paraphrase the expression without changing the meaning of the sentence. Ask the student to answer it as well as he or she can. If the student has trouble with a number of words or expressions or has another problem which cannot be quickly and easily rectified, simply indicate the problem on the front of the first page and thank the student.

- Occasionally, a student marks the answer to one or more items in the wrong place on the Questionnaire. The layout of the SDQ-II makes this mistake unlikely, but if it happens, simply tell the student to cross out the incorrect response and substitute the correct response. If this has occurred for a large number of responses, it may be necessary to transfer the correct responses to a new questionnaire.

Although this standard administration procedure works well in most settings, early experience indicated that the Questionnaire may cause problems for students who do not have an adequate command of written English. For these students, reading the items aloud may be helpful. This read-aloud procedure may also be useful when administering the SDQ-II to large groups where it is necessary that students finish in a reasonable amount of time or at approximately the same time. In these cases, use the following Read-Aloud Administration Procedure.

Read-Aloud Administration Procedure

This procedure is adapted from those used for the SDQ-I for which items are always read aloud.

- The read-aloud procedure begins the same as the standard procedure. The students are given copies of the Questionnaire and assured of the confidentiality of their responses. Toward the end of the instructions, after the statement, "Do not leave out any of the sentences," add the following statements:

  When we turn to the next page, I will read the sentences aloud. The reason for reading them aloud is to be sure that everyone spends the same amount of time on each item and that everyone finishes at the same time. We will be going quite fast, and you will have to mark your answer immediately. Then listen to the next sentence. If you fall behind, leave out the sentences you have not done. Listen to the sentence I am reading and answer that one. I will allow you time at the end to go back to any sentences that you have left out.

- When you are ready to begin, say: **Turn over the page and begin. Once you have started, PLEASE DO NOT TALK.** Be sure to stop any talking, commenting, and deliberate or unconscious vocalization.

- After the students have turned the page, begin reading the sentences in a clear, strong voice. Read the sentence number before the start of each sentence. The sentences should be read at a fairly rapid and steady pace (approximately eight sentences per minute). Read the sentence twice without any pause. Then pause briefly and begin reading the next sentence. Students may be surprised at how fast you are reading the sentences, but they will quickly keep pace.

- Do not stop to answer any questions once you have begun reading the sentences. If a student interrupts you during the administration of the items to ask the meaning of a word or the interpretation of an item, ask the student to wait until you have finished reading all the sentences. The student should be encouraged to continue with the other items and leave the problem item until the end.

  There may be a few students who do not keep pace with the administration, no matter how often they are encouraged to do so. If they persist after several reminders, it is best to allow them to proceed at their own pace. Allow such students time to complete the SDQ-II after all the sentences have been read aloud, and check to see that they have had no problems. Similarly, there may be students who want to go ahead of the administration, particularly if the pace of administration is not reasonably fast. Once again, encourage them to stay with the group, but allow them to proceed at their own pace if they persist.

- After you have completed reading all the sentences, say:

  Now I will give you a minute or two to go back to any sentences which you left out. Be sure you have *one*, and only one, answer for each sentence. Please do this now. When you have completed all the sentences, put your paper face up on your desk and wait quietly for the rest to finish. If there are any questions about completing the sentences, hold up your hand, and I will come to you.

- If there are any questions, go to the individual student. If a student has trouble understanding a few words or expressions, paraphrase the expression without changing the meaning of the sentence. Ask the student to answer it as well as he or she can. If the student has trouble with a number of words or expressions or has another problem which cannot be quickly and easily rectified, simply indicate the problem on the front of the first page and thank the student.

- Occasionally, a student marks the answer to one or more items in the wrong place on the Questionnaire. The layout of the SDQ-II makes this mistake unlikely, but if it happens, simply tell the student to cross out the incorrect response and substitute the correct response. If this has occurred for a large number of responses, it may be necessary to transfer the correct responses to a new questionnaire.

Scoring Instructions

Responses to the SDQ-II may be conveniently scored with the SDQ-II Scoring and Profile Booklet. The booklet provides for the calculation of the 11 individual scale raw scores and a Total Self-Concept raw score. Calculation of each of the scores is described below.

Calculating Individual Raw Scores

Complete the identifying information at the top of the SDQ-II Scoring and Profile Booklet by copying the information from the front of the student's SDQ-II Questionnaire.

		False	Mostly False	More False Than True	More True Than False	Mostly True	True	
1. Mathematics is one of my best subjects	1	☐	☐	☑	☐	☐	☐	1
2. Nobody thinks that I'm good looking	2	☐	☑	☐	☐	☐	☐	2
3. Overall, I have a lot to be proud of	3	☐	☐	☐	☐	☑	☐	3
4. I sometimes take things that belong to other people	4	☑	☐	☐	☐	☐	☐	4
5. I enjoy things like sports, gym, and dance	5	☐	☐	☐	☑	☐	☐	5
6. I'm hopeless in English classes	6	☑	☐	☐	☐	☐	☐	6
7. I am usually relaxed	7	☐	☐	☐	☑	☐	☐	7
8. My parents are usually unhappy or disappointed with what I do	8	☐	☑	☐	☐	☐	☐	8
9. People come to me for help in most school subjects	9	☐	☐	☐	☑	☐	☐	9
10. It is difficult to make friends with members of my own sex	10	☐	☐	☐	☑	☐	☐	10
11. People of the opposite sex whom I like don't like me	11	☐	☐	☑	☐	☐	☐	11
12. I often need help in mathematics	12	☐	☐	☑	☐	☐	☐	12
13. I have a nice looking face	13	☐	☐	☐	☐	☑	☐	13
14. Overall, I am no good	14	☑	☐	☐	☐	☐	☐	14
15. I am honest	15	☐	☐	☐	☐	☐	☑	15
16. I am lazy when it comes to things like sports and hard physical exercise	16	☐	☑	☐	☐	☐	☐	16
17. I look forward to English classes	17	☐	☐	☐	☐	☐	☑	17
18. I worry more than I need to	18	☐	☐	☐	☐	☐	☑	18
19. I get along well with my parents	19	☐	☐	☐	☐	☑	☐	19
20. I'm too stupid at school to get into a good university	20	☐	☑	☐	☐	☐	☐	20

Figure 12. Sample Page from the SDQ-II Questionnaire

Calculation of the individual scale and total raw scores is done on the Score Calculation and Summary page in the Scoring and Profile Booklet. First, score the individual scales. In the first section of the Score Calculation and Summary page, under the name of each scale, is a column of item numbers that compose the scale, blanks in which the student's item scores should be recorded, and item means (in parentheses). Using the student's Questionnaire, transfer the score for each of the student's responses in the appropriate space after the item number. If the item number is not preceded by an asterisk (*), simply convert the child's response to each item into one of the following scores:

False = 1	More True Than False = 4
Mostly False = 2	Mostly True = 5
More False Than True = 3	True = 6

If the item number is preceded by an asterisk (*), reverse the value of the student's response by subtracting it from 7. For example, if the student responded "False" to an asterisked item, the response value will be 7 − 1, or 6. Transferring item scores is facilitated by the layout of the Score Calculation and Summary page. When read across, rather than down each scale, the items are in numerical sequence. That is, the first item of the Math scale is Item 1, the first item of the Physical Appearance scale is Item 2, and so forth. When scoring the Same-Sex Relations and Opposite-Sex Relations scales, be sure to use the section appropriate to the sex of the student. For these two scales, the page includes separate sections for males and females.

Then, simply sum the scores in each column to arrive at the individual scale raw scores. For example, sum the scores for Items 1, 12, 23, 34, 45, 56, 67, 78, 89, and 96 to arrive at the raw score for the Math scale. Write the sum for each scale in the blanks labeled "Raw Scale Scores" that appear below the columns of item numbers.

If the student omits *five or fewer* responses on the Questionnaire, the mean response for the missing item should be substituted for the missing item score. Item means are given in parentheses following the blanks next to their respective item numbers on the Score Calculation and Summary page. *If six or more responses are missing, the responses should be either not scored at all or interpreted cautiously.* Moreover, if more than one response is missing on any individual scale, the results should be interpreted cautiously. When a scale contains an item mean, the sum of the items will not be a whole number; therefore, the sum should be rounded to the nearest whole number.

To avoid errors, make certain that each item has been translated into the correct score and has been written in the blank next to the correct item number. As each individual scale score is summed, it should be recalculated to avoid errors in addition. Also, note that for the Math, General Self, Honesty–Trustworthiness, Verbal, Emotional Stability, General School, and Same-Sex Relations scales, the lowest possible raw score is 10, and the highest possible raw score is 60. For the Physical Appearance, Physical Abilities, Parent Relations, and Opposite-Sex Relations scales, the lowest possible raw score is 8, and the highest possible raw score is 48. Check the raw score obtained for the individual scale to ensure that it lies within its upper and lower limits.

Calculating a Total Raw Score

The individual scale raw scores are used to calculate the Total Self-Concept raw score. To calculate Total Self-Concept, copy the individual scale raw scores into the appropriate blanks in the Total Self-Concept Raw Score section of the Score

Calculation and Summary page. Sum the individual scale raw scores and write the total in the blank labeled "Total Self-Concept."

Again, always double-check these calculations to avoid errors in copying or addition. Note that the lowest possible raw score for the Total Self-Concept is 102 and the highest possible raw score is 612. Check the Total Self-Concept score obtained to ensure that it falls within these bounds.

An example of a completed Score Calculation and Summary page from an SDQ-II Scoring and Profile Booklet is provided in Figure 13A.

Converting Raw Scores to Scaled Scores

Raw scores for the SDQ-II individual scales and Total Self-Concept may be converted to mid-interval percentile ranks and standard scores—in this case nonnormalized *T* scores. Percentile-rank scores are directly interpretable as the percentage of children in the normative sample who scored at or below the self-concept score for a particular child. *T* scores have a mean of 50 and a standard deviation of 10.

The tables needed to convert raw scores into percentiles and *T* scores are presented in Appendix A. A combined norm table, based on the total normative sample, is provided in Appendix A-1 and may be useful if a group comparison is required. However, because there are sex differences on several of the scales, normative comparisons are reported separately for males and females in Appendices A-2 and A-3, respectively.

To convert the individual scale raw scores, locate the raw score in the left or right columns of the appropriate norm table. Then find the column entries for the corresponding percentile and *T* score for each individual scale raw score. Enter the percentile and *T* score for each scale in the spaces provided at the bottom of the *T*-Score Profile page in the Scoring and Profile Booklet. For example, Appendix A-2 for males indicates that a raw score of 26 on the Physical Abilities scale corresponds to a percentile equivalent of 11 and a *T* score of 36—almost 1 1/2 standard deviations below the mean for this scale.

Appendix A-4 presents the percentile equivalents for the Total Self-Concept raw score. Raw score ranges are presented for the combined normative sample and for separate groups of males and females. To convert the Total Self-Concept raw score, locate the raw score range in the appropriate column of the table. Then find the percentile equivalent in the far left or right column that corresponds to the appropriate raw score range. Enter the percentile for the Total Self-Concept score in the space provided at the bottom of the *T*-Score Profile page.

Similarly, Appendix A-5 presents the *T* scores for the Total Self-Concept raw score. To convert the Total Self-Concept raw score, locate the raw score range in the appropriate column of the table and find the *T* score in the far left or right column that corresponds to the raw score range. Enter the *T* score for the Total Self-Concept score in the space provided at the bottom of the *T*-Score Profile page.

After the percentile and *T*-score equivalents for the individual scale scores and Total Self-Concept scores have been recorded, complete the profile by plotting each *T* score on the profile grid. An example of a completed *T*-score profile is presented in Figure 13B.

Score Calculation and Summary

INDIVIDUAL SCALE RAW SCORES: For each scale, write the value, from 1 to 6, of the student's response in the blank beside the item number. For items marked with an asterisk (*), reverse the value of the score by subtracting the value of the student's response from 7. (For example, if the student responded "False" to an asterisked item, the response value will be 7 - 1, or 6.)

Math			Physical Appearance			General Self			Honesty-Trustworthiness			Physical Abilities		
Item	Score	Mean	Item	Score	Mean	Item	Score	Mean	Item	Score	Mean	Item	Score	Mean
1	3	(3.44)	2	*5	(4.31)	3	5	(4.65)	4	*6	(4.96)	5	4	(4.97)
12	*4	(3.73)	13	5	(3.81)	14	*6	(5.02)	15	6	(4.61)	16	*5	(4.80)
23	3	(2.82)	24	*3	(3.35)	25	5	(4.82)	26	*6	(3.16)	27	5	(4.43)
34	*3	(4.36)	35	4	(3.83)	36	*5	(4.69)	37	6	(3.92)	38	*4	(4.73)
45	2	(2.76)	46	*4	(4.11)	47	5	(4.81)	48	*6	(4.38)	49	3	(3.76)
56	*4	(3.88)	57	4	(3.73)	58	*6	(4.59)	59	6	(4.79)	60	*5	(4.87)
67	4	(3.87)	68	*6	(4.47)	69	5	(4.92)	70	*6	(4.20)	71	3	(3.81)
78	*2	(3.82)	79	6	(3.57)	80	*5	(4.73)	81	6	(4.80)	82	5	(5.04)
89	4	(3.70)				90	5	(5.05)	91	*6	(4.49)			
96	*4	(3.55)				97	*6	(5.03)	98	6	(4.58)			
RAW SCALE SCORES	33			37			53			60			34	

Verbal			Emotional Stability			Parent Relations			General School		
Item	Score	Mean	Item	Score	Mean	Item	Score	Mean	Item	Score	Mean
6	*6	(4.48)	7	4	(4.46)	8	*5	(4.62)	9	4	(3.24)
17	6	(3.38)	18	*1	(3.37)	19	5	(5.03)	20	*5	(4.40)
28	*6	(4.58)	29	3	(3.62)	30	5	(4.54)	31	5	(4.04)
39	6	(4.26)	40	*5	(4.49)	41	5	(4.91)	42	*6	(4.53)
50	*6	(4.59)	51	*3	(4.07)	52	*4	(4.18)	53	5	(4.40)
61	6	(3.72)	62	*5	(4.26)	63	4	(4.67)	64	*5	(4.75)
72	*6	(4.22)	73	*4	(3.98)	74	5	(5.43)	75	5	(4.15)
83	6	(4.08)	84	*3	(3.83)	85	5	(5.30)	86	*5	(4.47)
92	*5	(3.81)	93	4	(4.34)				94	5	(4.35)
99	6	(4.21)	100	*2	(3.38)				101	5	(4.81)
RAW SCALE SCORES	59			34			38			50	

Males Only						Females Only					
Same-Sex Relations			Opposite-Sex Relations			Same-Sex Relations			Opposite-Sex Relations		
Item	Score	Mean	Item	Score	Mean	Item	Score	Mean	Item	Score	Mean
10	*	(5.01)	11	*	(4.17)	10	*3	(5.01)	11	4	(4.17)
21		(4.86)	22		(4.39)	22	4	(4.86)	21	5	(4.39)
32	*	(4.99)	33	*	(4.41)	32	*3	(4.99)	33	*4	(4.41)
43		(4.38)	44		(3.93)	44	4	(4.38)	43	4	(3.93)
54	*	(5.07)	55	*	(4.72)	55	3	(5.07)	54	4	(4.72)
65		(5.39)	66		(4.52)	65	6	(5.39)	66	4	(4.52)
76	*	(4.97)	77	*	(4.61)	77	*5	(4.97)	76	5	(4.61)
87		(3.91)	88		(3.63)	87	4	(3.91)	88	5	(3.63)
95	*	(4.82)				95	*3	(4.82)			
102		(5.03)				102	4	(5.03)			
RAW SCALE SCORES							39			35	

NOTE: Substitute item means (rounded to the nearest whole number) for missing values if five or fewer responses are left blank. If more than five responses on the entire Questionnaire or more than one response for an individual scale are left blank, the test should be interpreted cautiously.

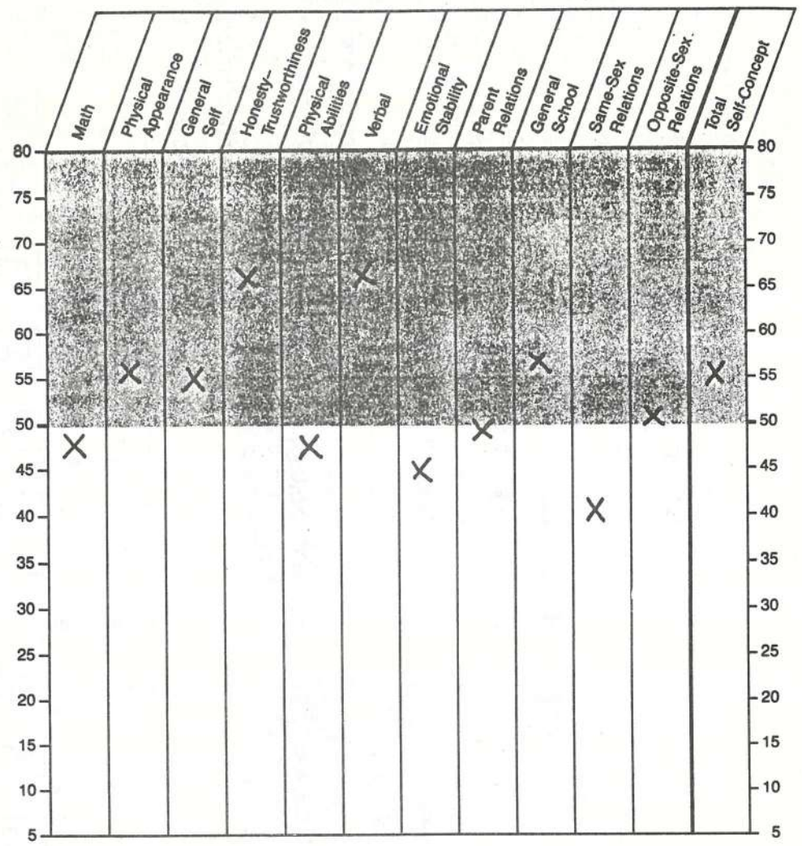
TOTAL SELF-CONCEPT RAW SCORE: Write the Individual Scale Raw Scores in the appropriate spaces below. Sum them and write the total in the space labeled "Total Self-Concept."

Math	33
+ Physical Appearance	37
+ General Self	53
+ Honesty-Trustworthiness	60
+ Physical Abilities	34
+ Verbal	59
+ Emotional Stability	34
+ Parent Relations	38
+ General School	50
+ Same-Sex	39
+ Opposite-Sex	35
= Total Self-Concept	472

Figure 13A. Calculating Individual and Total Self-Concept Raw Scores

T-Score Profile

Directions: Transfer the raw scores for the individual scales and Total Self-Concept score from page 2 to the spaces provided below the profile. Then, convert the raw scores to percentile ranks and *T* scores using the tables in Appendix A of the Manual. Record these values in the space provided and plot the *T* scores on the profile.



	Math	Physical Appearance	General Self	Honesty-Trustworthiness	Physical Abilities	Verbal	Emotional Stability	Parent Relations	General School	Same-Sex Relations	Opposite-Sex Relations	Total Self-Concept
Raw Scores:	33	37	53	60	34	59	34	38	50	39	35	472
Percentiles:	41	68	62	99	34	98	27	36	72	13	49	67
T Scores:	48	56	55	66	47	67	44	49	57	40	51	55

Note: *T* scores falling in the shaded area (i.e. *T* scores of 50 or above) represent above average self-concept; however, because of the skewed distribution of the scores, *T* scores above 50 are not readily interpretable.

Figure 13B. Plotting a *T*-Score Profile and Percentile–Rank Equivalents

Factor Scores

Factor analysis has played an important role in the development of the SDQ-II, and much of the published research described later is based on factor scores used to represent the 11 SDQ-II facets. These factor scores are based on analytically weighted responses rather than on the unweighted sums used to determine scale scores. Although factor scores and the unweighted scale scores are substantially correlated, factor analytically derived scores distinguish better among the SDQ-II facets than do the unweighted scale scores, and they are more clearly related to various criterion measures. It is anticipated that the SDQ-II will often be used in large-scale research projects where the user has ready access to computers. In such instances, users are encouraged to use factor scores derived from factor analyses rather than the results of hand-scored scales, particularly when the focus of research is to distinguish between different facets of self-concept. Factor scores based on the factor analysis of the normative sample of SDQ-II responses can be computed on the basis of the regressed factor scores presented in Appendix B. Because this factor analysis is based on such a large sample and provides such good support for the SDQ-II factor structure, these factor scores are recommended for research purposes.

Chapter 3.
Norms Development and Interpretation

This chapter describes the process of scale development and standardization and presents guidelines for interpreting the SDQ-II.

Scale Development

After several early revisions of the instrument, a pool of 140 items was constructed to represent the 11 facets of self-concept measured by the SDQ-II. This earlier version of the instrument was intentionally constructed to be longer than the current instrument so that the best possible items could be selected for subsequent revisions. The selection of a 122 of these items to represent best the 11 scales was substantially based on the first large-scale study with SDQ-II (Marsh, Parker, & Barnes, 1985). The final revision of the SDQ-II is based on the entire set of 5,494 responses in the normative sample. The selection of the 102 items for retention in SDQ-II was based on this sample. Each of the 11 scales is represented by 8 or 10 items, half of which are negatively worded. Items were selected on the basis of both item and factor analyses. The final item selection was designed to reduce the overall length of the total instrument and at the same time, to maintain appropriate levels of subscale reliability.

Standardization

Sample Description

The norms in this Manual are based on responses by 5,494 students (2,658 males and 2,836 females) from schools in metropolitan Sydney, Australia. Data were collected by the author and his colleagues in research studies described in subsequent chapters. For these studies the SDQ-II was administered to all students in intact classrooms during regular school hours. The numbers of subjects are reasonably balanced in terms of sex and year of school through grade (or year) 10. Historically, many Australian students leave school at the end of year 10; thus, there are fewer responses by students in years 11 and 12, and this sample may be less representative than the sample for years 7 through 10. Although no overall sampling plan was employed, care was taken to ensure that the selected schools were broadly representative of the population of school children in Sydney, Australia. The sample included schools from diverse regions of greater metropolitan Sydney; schools in working-class, middle-class, and upper-middle-class areas; single-sex and coeducational schools; and both public and Catholic schools. In general, students were not grouped according to ability levels in these schools. These norms may be less appropriate for students in other parts of the world, or for students who complete the instrument for other purposes or under different motivational circum-

stances. However, cross-cultural data presented in the SDQ-I Manual (Marsh, 1988) provide preliminary support for using this large, well-stratified sample for students outside Australia.

Derivation of Norms

Norms for the SDQ-II are presented as mid-interval percentile ranks and nonnormalized *T* scores. Nonnormalized *T* scores are similar to those used on the *Minnesota Multiphasic Personality Inventory* (MMPI; Hathaway & McKinley, 1951) and other widely used instruments. The *T*-score transformation uses the mean and standard deviation of the normative group (raw score minus the mean, divided by the standard deviation resulting in a *z* score, converted to a standard score with a mean of 50 and a standard deviation (*SD*) of 10). This transformation makes each of the scales comparable, in terms of standard deviation units, so that *T* scores can be entered on a profile to assess an individual's pattern of scores.

The user should note that the *T*-score transformation used in SDQ-II does not change the essential shape of the distributions of scale scores. This issue is discussed in depth in the SDQ-I Manual. Inspection of the shape of the raw-score distributions indicates that they are negatively skewed for nearly all SDQ-II scales. Skewness ranges from $^{-}0.10$ (Math) to $^{-}1.29$ (Parent Relations) with a median of $^{-}0.62$ for the total normative group, a range indicating that there is a higher proportion of scores above the mean (i.e., a large number of high scores) for most of the SDQ-II scales. Thus the user should keep in mind that 40*T* and 60*T* may not be equidistant from the mean in a percentile sense for the individual scales.

The skewness of the SDQ-II score distributions has a substantive explanation. The author suggests that the self-concepts found in samples of students in normative samples may be truly skewed; that is, most students tend to feel positively about themselves, and only rarely does a student feel very negatively about himself or herself. The SDQ-II scale with the most extreme skewness is the Parent Relations scale. Although some of this skew may be due in part to the idiosyncratic wording of the items in the Parent Relations scale, it is also the scale in which students' self-concepts are most uniformly positive. If the underlying distributions of self-concepts are skewed, a score transformation that preserves the shape of the distributions, such as the nonnormalized *T* score, is the method of choice for interpreting the scores.

Interpretation

The comparison of different scales for the same student provides a special set of problems in the appropriate scaling of SDQ-II scores. If all the scales of the SDQ-II were normally distributed with the classic bell-shaped curve, the comparison between scores would be direct. Each score could be entered on a profile sheet without concern for the differences among scale distributions. For example, the student's Math self-concept score could be directly compared to his or her Verbal self-concept score. However, because each of the SDQ-II scales has a skewed distribution of varying degrees, the issue of scaling and profile comparability is more complicated. Two approaches to scale-score comparison are recommended, but with certain cautions: the comparison of percentiles and the comparison of standard *T* scores. An alternative approach, the comparison of raw scores, is not recommended.

Raw Scores. The simple comparison of raw scores on a test or inventory is rarely defensible. On achievement tests, a subtest such as reading comprehension may be more difficult on the average than a subtest such as vocabulary. On personality inventories, minor wording changes in a few items from one scale to another could systematically raise or lower all scores on one of the scales. An exception is the academic scales of SDQ-I, which have strictly parallel wording of all items (e.g., "I hate Math," "I hate Reading"). However, the SDQ-II academic scales do not have a complete set of parallel items. For this reason differences between two raw scale scores on the SDQ-II generally should not be used to infer underlying self-concept differences. As indicated in Chapter 4, the standard errors of measurement (SE_Ms) of the raw scores range from 2.9 (Opposite-Sex Relations) to 4.2 (Emotional Stability and Math); thus, differences smaller than these should never be interpreted as significant. The Total Self-Concept Score has a standard error of measurement of 16.6.

Percentiles. When raw scores are converted to percentiles, the student's relative position in the standardization sample can be identified. The percentile shows the percentage of students in the normative sample whose scores are equal to or less than the student's score.

Because half of the normative sample had self-concept scores that fell between the 25th and 75th percentiles, scores in this range are neither particularly high nor particularly low. In addition, small differences, particularly at the high end of the scale, can result in large percentile differences. For these reasons, interpretation at the high end of the SDQ-II scales is discouraged. Even in the middle range one can only assume that the student has average self-concept. Therefore, only scores at the low end of the scale are readily interpretable and diagnostically meaningful. For example, if a student is in the average range (e.g., 40th percentile) for Math but very low in Verbal (e.g., 8th percentile), the difference can readily be interpreted as a deficit in Verbal self-concept.

***T* Scores.** The use of *T* scores makes possible the construction of a profile for the individual student (see the SDQ-II Scoring and Profile Booklet). When raw scores are transformed into *T* scores, the resulting scores are comparable on the basis of standard deviation units. A *T* score of 30, two standard deviations below the mean, would have the same meaning on each of two scales. For example, if both Physical Abilities and Physical Appearance were 30*T*, the student is two standard deviations below the mean of the normative sample on both scales—with both being remarkably low.

As described in Chapter 4, the standard errors of measurement of the *T* scores range from 3.0 (Physical Appearance) to 4.1 (Emotional Stability); the standard error is 2.4 for the Total Self-Concept score. Differences smaller than these should never be considered significant. Also, a common rule of thumb is that the "true" score of an individual lies in the interval created by ± 1 standard error.

Thus, individual scale *T* scores would have intervals around the true score of about 8 points in width into which the observed *T* score may fall. However, comparison of the different individual scale scores should be made cautiously because the skews for each scale-score distribution vary substantially in some instances, particularly between the Parent Relations scale (high negative skewness) and the Math scale (low negative skewness). Thus, when *T* scores for two different scales are the same, the corresponding percentile ranks will not be the same unless the distribution of scores is similar. Because most of the SDQ-II scales are only moderately skewed, the comparison of specific scale scores may be reasonable.

Individual Scales

SDQ-II research has focused on the specific SDQ-II facets rather than on the General Self or Total Self-Concept scores, and, therefore, these individual scales

should be the primary bases for interpreting the responses to the SDQ-II. Each of these scales has high face validity. (See Figures 1 through 11, pages 3–6, for the inferred meaning of the 11 SDQ-II scales and the items composing each scale.) In general, a high scale score indicates that the respondent has a positive self-perspective in that area, whereas a low score indicates a negative self-perspective. However, two special situations require further consideration.

First, a respondent may have a self-perspective that is quite unrealistic when compared to objective information. For example, a person may be poor at mathematics and yet have a positive Math self-concept. However, self-concept is defined according to how a person actually views himself or herself and not according to how a person *should* view himself or herself. Insofar as the person has responded honestly, his or her responses reflect a valid inference about self-concept even if those responses are unrealistic. Hence, when external indicators or the opinions of observers differ from the student's perceptions, it is the external sources that probably lack validity as measures of the student's self-concept.

Second, interpretations of the SDQ-II are based on the assumption that the subject is responding honestly. Because the SDQ-II items and scales are straightforward, it would be easy to respond so as to give a "good" or a "poor" impression. The strength of the psychometric properties of the SDQ-II—particularly the factor analyses and relations to external criteria—indicates that this kind of responding is not typical in SDQ-II research. However, SDQ-II research has nearly always been conducted in a setting where subjects have little external incentive to respond dishonestly. In a setting where respondents are externally motivated to look "good" or "bad," the responses must always be interpreted cautiously, as is the case with all self-evaluation and self-report instruments.

Two approaches typically are taken in this situation to insure valid test results at the level of an individual student. The first, which is recommended for use with the SDQ-II, is to assure respondents that their responses will remain confidential and that responding candidly will not harm them and may even be helpful. The second, sometimes used with other self-report instruments, is to construct a separate scale to measure some construct related to "social desirability responding." However, responding positively to socially desirable attributes is also the basis of self-concept inferences; consequently such an approach is untenable in self-concept research (see the SDQ-I Manual for further discussion).

If the user is reasonably assured that the student has given an accurate and candid view of himself or herself, the profile of individual scores can be inspected for patterns of high and low scores. There are no absolute cutoff points that define high and low self-concept, and interpretations should be made in relation to percentile ranks and *T* scores derived from the norm tables. Several general guidelines, however, may be useful. First, as discussed in Chapter 4, the standard errors of measurement for the SDQ-II individual scales range from approximately 3 to 4 *T*-score points. Thus, smaller differences between scales should not be interpreted. Second, as mentioned earlier, the SDQ-II scales are skewed and typically have higher proportions of scores above the mean.

Thus, differences among high scores should be interpreted with great caution because small response differences at the high end of the scales can result in relatively large differences in percentile ranks. However, very low scores such as those below 30*T* should be noted because they are atypical and may signal an area of individual concern. Finally, it is expected that most students will have a range of self-concepts in different areas. Thus, even the most able student may have average or even below-average self-concepts in some areas. If the below-average self-concept score is not extremely low (e.g., 30*T* or less), the student may simply be reporting a realistic view of his or her attributes in that area. However, if the student consistently has very high or very low self-concept scores across all areas, further investigation into the reasons for this unusual pattern is recommended.

One method of examining unusual patterns is to study the actual item responses of an individual. After responses have been recorded on the Score Calculation and Summary page, the pattern of responses should be examined to determine if the respondent has answered the questions appropriately (i.e., not randomly or carelessly). A comparison of positively and negatively worded items may show that the respondent was not paying attention to item content. Also, the original responses on the questionnaire may indicate whether the respondent has given the same response to all items or has used some pattern of invalid responses.

General Self Scale

Historically, self-concept research has emphasized a general self-concept instead of specific facets of the self. However, this general self was typically inferred from responses to a hodgepodge of items that were not selected or balanced with respect to any theoretically defensible components of the self. Consequently, measures of general self were often idiosyncratic to a specific instrument (see Marsh & Smith, 1982). Alternatively, other researchers constructed relatively unidimensional scales that were intended to measure a superordinate construct (e.g., "general self-esteem"). The most well-known of this type of scale is the Rosenberg (1965) scale which was used as the conceptual basis of the General Self scale on the SDQ instruments (also see Harter, 1982). The General Self scale on the SDQ-II, as on the Rosenberg scale, implies a general or overall positive self-perspective that is not specific to any particular facet of self-concept but could be applied to each specific facet of the self. Thus, the user should examine the General Self scale separately from the other individual scales and note any extremely low scores on this global self-evaluative scale. Interpretations of general self-concept and its theoretical and empirical bases are discussed further in Chapter 5 (see also Marsh, 1986b).

Total Self-Concept Score

The Total Self-Concept score on the SDQ-II represents the sum of the 11 individual scales. Thus, it is an aggregate of many distinct facets of self-concept and should be interpreted with caution. It is most highly correlated with the General Self scale, which measures a global evaluation of self ("self-esteem"). The Total Self-Concept score is intended to provide a global measure like those traditionally measured in many other self-concept instruments. Also, the original justification for a total score was based on the hierarchical Shavelson et al. model (1976; see Chapter 5) in which the apex of total self-concept is divided into academic and nonacademic facets, each of which has numerous individual facets. Subsequent research (see Chapter 5) has demonstrated that the hierarchy may be weaker than originally hypothesized and that Math and Verbal factors are particularly distinct. Thus, the SDQ-II does not include a total academic score such as that found in the SDQ-I. Also, the nonacademic factors of the SDQ-II appear to be too diverse to allow a total nonacademic score to be included. For these reasons, users are encouraged to emphasize the specific facets of self-concept rather than the Total Self-Concept score.

This chapter has described the process of scale development and standardization and has presented guidelines for interpreting SDQ-II scale scores for the various facets as well as at the more global levels of the General Self and Total Self-Concept scores. Chapter 4 presents evidence of test reliability and stability over time.

Chapter 4. Reliability

Internal consistency and stability are important psychometric properties of any measurement instrument. This chapter examines SDQ-II responses in relation to these two types of reliability estimates.

Internal Consistency

Reliability refers to the extent to which responses are due to systematic sources of variance. The primary basis for estimating reliability in SDQ-II research has been the internal consistency of item responses in each of the SDQ-II scales. Coefficient alpha estimates of reliability have been presented in many of the published studies listed in the references of this Manual. For purposes of this Manual, coefficient alphas were computed from the total normative sample (see Table 2). Internal consistency for the Total Self-Concept score was .94. Coefficient alphas for the 11 facet scores vary from .83 for Emotional Stability to .91 for Physical Appearance (median = .86).

Table 2. Internal Consistency and Standard Errors of Measurement (SE_Ms)

Scale	Coefficient Alpha	SE_M Raw Score	SE_M *T* Score
Physical Abilities	.85	3.6	3.9
Physical Appearance	.91	3.1	3.0
Opposite-Sex Relations	.90	2.9	3.2
Same-Sex Relations	.86	3.1	3.8
Parent Relations	.87	3.2	3.6
Honesty–Trustworthiness	.84	4.1	4.0
Emotional Stability	.83	4.2	4.1
Math	.90	4.2	3.2
Verbal	.86	4.1	3.7
General School	.87	3.7	3.6
General Self	.88	3.4	3.5
Total Self-Concept Score	.94	16.6	2.4

Note: N = 5,494. Based on the formula $SE_M = SD\sqrt{1 - r_{xx}}$ with the raw score standard deviations and alpha coefficients (r_{xx}) based on the combined standardization sample. Differences between the standardized and unstandardized item scores did not exceed .01; therefore, correlation coefficients based on only the standardized item scores are reported.

Other evidence of internal consistency comes from the intercorrelations among items within each scale. Correlations among responses to the 8 or 10 items designed to measure each SDQ-II scale are presented in Table 3, along with the item means and standard deviations (these item means may be used for replacing omitted responses; see Chapter 2). For each item, the corrected item–scale correlation (i.e., the correlation between an item and the sum of responses to other items in the same scale) is also presented in Table 3. The corrected item–scale correlations vary from .35 to .80 (median = .61). These summary statistics further demonstrate that each individual item is significantly and substantially correlated with the other items designed to measure the same facet of self-concept.

Standard Error of Measurement

Estimates of the standard error of measurement for each of the SDQ-II scores are shown in Table 2. These estimates are based on the alpha coefficients in Table 2, the standard deviations of raw scores in the normative sample (see Appendix A-1), and the fact that *T* scores have a standard deviation of 10. The results show that raw-score standard errors of measurement vary from 2.9 to 4.2 for individual scales, and the standard error of measurement for the Total Self-Concept score is 16.6. *T*-score standard errors of measurement vary from 3.0 to 4.1 for individual scales; that for the Total Self-Concept score is 2.4. The practical implications of these values are discussed in Chapter 3. The usual guideline is that the observed score of an individual should exist in an interval of ±1 standard error around the individual's "true" score (a hypothetical, errorless score). Therefore, the user should ignore scale differences smaller than 3 or 4 *T*-score points and should recognize that the individual's true score may be as much as 8 points different from the observed score computed on the SDQ-II. Sources of error include fatigue, mood changes, events in the individual's environment, atypical methods of test administration, and the like.

Stability of SDQ-II Responses Over Time

Marsh and Peart (1988; also see Chapter 8) examined responses by a group of 137 high school girls before and after an intervention designed to enhance physical fitness. Coefficient alpha estimates of reliability based on SDQ-II responses before and after the intervention (see Table 4) compare favorably with those based on the entire SDQ-II normative sample (see Table 2). Despite the effects of the intervention that took place between these two data collections, short-term stability coefficients are substantial, varying between .72 for Emotional Stability and .88 for Math (median = .79).

In summary, these findings demonstrate that SDQ-II responses are internally consistent and stable over time. Given the factor analytic results summarized in Chapter 5 and additional support for construct validity discussed in subsequent chapters, these results provide strong support for the psychometric properties of SDQ-II responses.

Item Number	Mean	*SD*	Corrected Item–Total Correlation	Correlations Among Items									
Physical Abilities				5	16	27	38	49	60	71	82		
5	4.97	1.54	.66	—									
16	4.80	1.59	.62	.47	—								
27	4.43	1.65	.73	.60	.49	—							
38	4.73	1.63	.60	.45	.45	.56	—						
49	3.76	1.74	.60	.43	.39	.61	.45	—					
60	4.87	1.68	.52	.45	.51	.40	.37	.28	—				
71	3.81	1.81	.48	.29	.38	.43	.35	.46	.26	—			
82	5.04	1.43	.56	.59	.39	.45	.38	.34	.40	.25	—		
Physical Appearance				2	13	24	35	46	57	68	79		
2	4.31	1.59	.63	—									
13	3.81	1.66	.76	.49	—								
24	3.35	1.72	.61	.47	.48	—							
35	3.83	1.66	.80	.53	.77	.52	—						
46	4.11	1.74	.69	.55	.53	.51	.56	—					
57	3.73	1.62	.70	.51	.69	.44	.70	.46	—				
68	4.47	1.68	.71	.53	.59	.49	.61	.68	.50	—			
79	3.57	1.74	.65	.43	.56	.46	.62	.49	.53	.50	—		
Opposite-Sex Relations				11	21/22	33	43/44	54/55	66	76/77	88		
11	4.17	1.67	.55	—									
21/22	4.39	1.56	.76	.43	—								
33	4.41	1.46	.69	.43	.57	—							
43/44	3.93	1.62	.72	.44	.61	.56	—						
54/55	4.72	1.51	.72	.45	.67	.56	.55	—					
66	4.52	1.49	.67	.36	.61	.59	.55	.55	—				
76/77	4.61	1.48	.65	.49	.54	.49	.52	.57	.43	—			
88	3.63	1.67	.70	.44	.61	.54	.64	.52	.54	.49	—		
(Note: Four of the items depend on whether the respondent is a boy or a girl.)													
Same-Sex Relations				10	21/22	32	43/44	54/55	65	76/77	87	95	102
10	5.01	1.35	.64	—									
21/22	4.86	1.29	.64	.47	—								
32	4.99	1.35	.62	.52	.44	—							
43/44	4.38	1.51	.49	.32	.43	.32	—						
54/55	5.07	1.30	.62	.46	.46	.46	.39	—					
65	5.39	1.11	.51	.37	.38	.38	.24	.34	—				
76/77	4.97	1.38	.54	.40	.39	.43	.36	.45	.30	—			
87	3.91	1.35	.61	.47	.52	.43	.34	.44	.41	.35	—		
95	4.82	1.61	.49	.43	.33	.43	.27	.37	.28	.33	.34	—	
102	5.03	1.24	.49	.33	.41	.30	.33	.33	.44	.27	.38	.24	—
(Note: Four of the items depend on whether the respondent is a boy or a girl.)													

Note: N = 5,494; r_{it} = corrected item–total correlations, the correlations between each item and the sum of the remaining items in the scale.

Table 3. Item Statistics by Scale *(continued)*

Item Number	Mean	*SD*	Corrected Item–Total Correlation	Correlations Among Items									
Parent Relations				**8**	**19**	**30**	**41**	**52**	**63**	**74**	**85**		
8	4.62	1.68	.56	—									
19	5.03	1.39	.76	.45	—								
30	4.54	1.78	.60	.45	.51	—							
41	4.91	1.50	.65	.39	.60	.42	—						
52	4.18	1.82	.59	.46	.52	.46	.42	—					
63	4.67	1.58	.64	.37	.60	.45	.56	.43	—				
74	5.43	1.20	.64	.41	.62	.42	.48	.41	.43	—			
85	5.30	1.27	.60	.38	.60	.38	.48	.34	.44	.54	—		
Honesty – Trustworthiness				4	15	26	37	48	59	70	81	91	98
4	4.96	1.58	.53	—									
15	4.61	1.41	.64	.37	—								
26	3.16	1.74	.53	.34	.36	—							
37	3.92	1.57	.55	.29	.49	.40	—						
48	4.38	1.94	.53	.38	.35	.34	.25	—					
59	4.79	1.50	.54	.33	.48	.27	.36	.35	—				
70	4.20	1.76	.61	.41	.40	.43	.34	.60	.33	—			
81	4.80	1.37	.51	.31	.48	.26	.34	.30	.40	.32	—		
91	4.49	1.61	.62	.46	.46	.47	.48	.34	.37	.41	.34	—	
98	4.58	1.34	.35	.17	.34	.18	.28	.18	.28	.21	.27	.24	—
Emotional Stability				7	18	29	40	51	62	73	84	93	100
7	4.46	1.57	.51	—									
18	3.37	1.85	.57	.28	—								
29	3.62	1.86	.53	.30	.35	—							
40	4.49	1.60	.49	.33	.30	.24	—						
51	4.07	1.73	.44	.29	.27	.42	.23	—					
62	4.26	1.75	.56	.41	.44	.30	.33	.26	—				
73	3.98	1.75	.52	.32	.36	.25	.47	.20	.36	—			
84	3.83	1.83	.63	.31	.46	.62	.33	.41	.40	.36	—		
93	4.34	1.70	.36	.38	.19	.24	.25	.20	.23	.22	.24	—	
100	3.38	1.84	.60	.34	.57	.31	.36	.27	.45	.47	.41	.20	—
Math				1	12	23	34	45	56	67	78	89	96
1	3.44	1.95	.77	—									
12	3.73	1.83	.59	.46	—								
23	2.82	1.81	.60	.57	.26	—							
34	4.36	1.64	.60	.47	.56	.29	—						
45	2.76	1.77	.60	.56	.27	.60	.30	—					
56	3.88	1.80	.65	.52	.57	.33	.53	.34	—				
67	3.87	1.74	.74	.65	.57	.40	.55	.44	.64	—			
78	3.82	1.93	.65	.55	.36	.49	.44	.50	.44	.47	—		
89	3.70	1.77	.69	.62	.51	.39	.51	.43	.55	.71	.44	—	
96	3.55	1.98	.65	.60	.33	.64	.37	.53	.40	.43	.58	.41	—

Table 3. Item Statistics by Scale *(continued)*

Item Number	Mean	*SD*	Corrected Item–Total Correlation	Correlations Among Items									
Verbal				**6**	**17**	**28**	**39**	**50**	**61**	**72**	**83**	**92**	**99**
6	4.48	1.60	.67	—									
17	3.38	1.81	.51	.42	—								
28	4.58	1.62	.55	.42	.18	—							
39	4.26	1.56	.58	.46	.41	.31	—						
50	4.59	1.71	.57	.41	.21	.61	.30	—					
61	3.72	1.65	.60	.47	.58	.27	.44	.30	—				
72	4.22	1.91	.44	.32	.26	.40	.19	.50	.31	—			
83	4.08	1.62	.67	.58	.43	.39	.55	.37	.51	.27	—		
92	3.81	1.82	.41	.31	.19	.38	.29	.32	.25	.18	.30	—	
99	4.21	1.61	.70	.58	.42	.38	.56	.39	.48	.27	.62	.31	—
General School				9	20	31	42	53	64	75	86	94	101
9	3.24	1.66	.43	—									
20	4.40	1.77	.62	.30	—								
31	4.04	1.68	.47	.32	.35	—							
42	4.53	1.26	.60	.26	.44	.26	—						
53	4.40	1.48	.60	.34	.42	.35	.35	—					
64	4.75	1.54	.59	.22	.50	.29	.43	.41	—				
75	4.15	1.55	.68	.40	.44	.39	.44	.52	.42	—			
86	4.47	1.30	.59	.23	.43	.25	.25	.67	.35	.43	—		
94	4.35	1.53	.67	.38	.46	.41	.42	.52	.44	.64	.40	—	
101	4.81	1.40	.57	.21	.43	.26	.45	.38	.50	.41	.48	.40	—
General Self				3	14	25	36	47	58	69	80	90	97
3	4.65	1.44	.65	—									
14	5.02	1.42	.60	.41	—								
25	4.82	1.26	.64	.53	.36	—							
36	4.69	1.50	.62	.39	.45	.43	—						
47	4.81	1.26	.62	.51	.37	.58	.45	—					
58	4.59	1.62	.63	.54	.46	.40	.48	.37	—				
69	4.92	1.29	.55	.46	.33	.47	.36	.44	.34	—			
80	4.73	1.58	.61	.42	.47	.36	.50	.38	.52	.34	—		
90	5.05	1.24	.49	.42	.27	.50	.30	.44	.28	.43	.27	—	
97	5.03	1.40	.64	.40	.61	.40	.48	.38	.50	.25	.52	.28	—

Note: $N = 5{,}494$; r_{it} = corrected item–total correlations, the correlations between each item and the sum of the remaining items in the scale.

Table 4. Internal Consistency and Test–Retest Reliability Over a 7-Week Interval

Variable	Coefficient Alpha r_1	Coefficient Alpha r_2	Test–Retest Reliability r_{12}
Physical Abilities	.89	.90	.86
Physical Appearance	.91	.92	.75
Opposite-Sex Relations	.93	.95	.79
Same-Sex Relations	.90	.93	.76
Parent Relations	.90	.91	.77
Honesty–Trustworthiness	.77	.76	.73
Emotional Stability	.87	.90	.72
Math	.92	.92	.88
Verbal	.88	.90	.85
General School	.89	.92	.82
General Self	.91	.95	.85

Note: N = 137. All values significant at the .01 level. r_1 = alpha coefficient of initial test scores; r_2 = alpha coefficient of retest scores; r_{12} = initial test–retest reliability (i.e., stability coefficients).

From "Competitive and Cooperative Physical Fitness Training Programs for Girls: Effects on Physical Fitness and on Multidimensional Self-Concepts" by H. W. Marsh and N. D. Peart, 1988, *Journal of Sport Psychology, 10*, p. 399. Copyright 1988 by Human Kinetics Publishers. Adapted by permission of the publisher.

Chapter 5. Construct Validity

Construct validity refers to the degree to which scores derived from an instrument function empirically as indicators of a hypothetical construct or a theoretical trait. Because self-concept is a hypothetical construct *and* a theoretical trait, construct validation is central to assessing the validity of self-concept responses. Construct validity is multiply determined; there is no single, simple criterion of what a particular test is supposed to measure. Construct validity requires the gradual accumulation of data from a variety of sources across many different studies. As noted by Anastasi (1976), "Any data throwing light on the nature of the trait under consideration and conditions affecting its development and manifestations are grist for this validity mill" (p. 151). Recognizing that there are many approaches to construct validation, Anastasi specifically mentioned factor analysis, multitrait–multimethod analysis, developmental changes, relations with other tests, and the effects of experimental interventions. Each of these aspects of construct validity has been addressed in SDQ research.

The initial problem in construct validation is defining the hypothetical construct or theoretical trait to be measured. Self-concept, like many psychological constructs, suffers in that "everybody knows what it is," and researchers do not feel compelled to provide any theoretical definition of what they are measuring. Shavelson et al. (1976) reviewed both theoretical and empirical research and used this review as the basis for developing a theoretical model of self-concept. The Shavelson et al. model was the initial basis for the design of the SDQ instruments and for subsequent SDQ research. An important postulate of the Shavelson et al. model is that self-concept is multidimensional. In a review of research based on the Shavelson et al. model, Marsh and Shavelson (1985) concluded that self-concept cannot be adequately understood if its multidimensionality is ignored.

Construct validity investigations can be classified as within-network or between-network studies. Within-network studies explore the multidimensionality of self-concept and attempt to show that it has consistent, distinct components. These studies typically employ factor analysis or, perhaps, multitrait–multimethod approaches. Within-network studies of SDQ-II responses are summarized primarily in the present chapter. Between-network studies attempt to demonstrate a theoretically consistent—or at least logical—pattern of relations between measures of self-concept and other constructs. These studies relate self-concept responses to other variables such as age and sex effects (Chapter 6), measures of academic performance (Chapter 7), the effects of experimental interventions designed to enhance self-concept (Chapter 8), other self-related constructs (Chapter 9), and a host of other variables (see also the SDQ-I and SDQ-III manuals). Between-network studies of SDQ-II responses are summarized in Chapters 6 through 9.

An obvious principle underlying SDQ research is that at least partial resolution of within-network issues must logically precede between-network research. Unless there exist a theoretical definition of a construct, a reasonable way to measure a construct, and empirical support for the facets posited to compose a multidimensional construct, then it makes little sense to pursue between-network studies. Adherence to this principle distinguishes the present approach from most other

self-concept research, particularly research conducted prior to the 1980s. Consistent with this principle, this chapter focuses on the theoretical and empirical support for the dimensionality of SDQ-II responses.

The multidimensionality of self-concept is a basic assumption of the Shavelson et al. model, which has important implications for understanding self-concept. Thus, for example, a person may have a positive self-concept in Opposite-Sex Relations but still have a poor Math self-concept. The primary purpose of this chapter is to provide the theoretical and empirical justification for this assumption. The theoretical development of the SDQ-II began with the Shavelson et al. model briefly described below. The factors included in the SDQ-II and those of the SDQ-I and SDQ-III were based in large part on the Shavelson et al. model. Item pools for each factor were initially developed, and preliminary analyses were conducted to refine the items designed to measure each scale. Once suitable versions of the SDQ had been developed, additional research was conducted to examine psychometric characteristics of the instruments, to relate the various scales to other constructs, and to test assumptions of the Shavelson et al. model. Hence, the development of the SDQ represents an interplay between theory and empirical research, an interplay which is the basic ingredient of construct validation.

The Shavelson et al. Model

Reviews of self-concept research continue to identify shortcomings no less dramatic than the lack of a theoretical basis for defining and interpreting the construct and the poor quality of instrumentation used to measure it. In an attempt to remedy this situation, Shavelson et al. (1976) reviewed existing research and self-concept instruments and developed a multifaceted, hierarchical model of self-concept. Self-concept, broadly defined by Shavelson et al., is a person's perceptions regarding himself or herself. These perceptions are formed through experience with and interpretations of one's environment. They are influenced especially by evaluations by significant others, by reinforcements, and by attributions of one's own behavior. The Shavelson et al. model and its relation to SDQ research is reviewed in detail in the SDQ-I and SDQ-III manuals.

Shavelson et al. presented one possible representation of the hierarchical model in which general self-concept appears at the apex and is divided into academic and nonacademic self-concepts at the next level. Academic self-concept is further subdivided into self-concepts in particular subject areas (e.g., English, mathematics). Nonacademic self-concept is subdivided into three areas: social self-concept (which is divided into relations with peers and with significant others), emotional self-concept, and physical self-concept (which is divided into physical ability and physical appearance). Further levels of division are hypothesized for each of these specific self-concepts so that at the base of the hierarchy, self-concepts are quite specific and closely related to actual behavior. This model posits a structure of self-concept that resembles a hierarchical model of intellectual abilities in which general self-concept (such as Spearman's *g*) is at the apex and is divided into two components which are then divided into group and specific factors.

The self-concept facets proposed in the Shavelson et al. model, as well as their hypothesized structure, were heuristic and plausible but had not been empirically validated by Shavelson et al. or in any of the research discussed in their review. Through the mid-1970s self-concept instruments typically consisted of a hodgepodge of self-referent items, and "blind" applications of exploratory factor analyses usually failed to identify salient, replicable facets (Marsh & Smith, 1982). More recently, researchers have developed self-concept instruments to measure specific

facets that are at least loosely based on an explicit theoretical model, and then have used factor analysis to support these facets (e.g., Boersma & Chapman, 1979; Dusek & Flaherty, 1981; Fleming & Courtney, 1984; Harter, 1982; Soares & Soares, 1982; also see research summarized in the SDQ-I Manual). Recent reviews of this research (Byrne, 1984; Marsh & Shavelson, 1985; Shavelson & Marsh, 1986) support the multifaceted structure of self-concept and indicate that self-concept cannot be adequately understood if its multidimensional character is ignored. In her review of self-concept models, Byrne concludes:

> Although no one model to date has been sufficiently supported empirically so as to lay sole claim to the within-network structure of the construct, many recent studies, in particular those of Marsh and his colleagues, are providing increasingly stronger support for the hierarchical model. (p. 449)

Factor Analytic Studies

The theoretical justification for the 11 SDQ-II factors and their relation to the Shavelson et al. model are discussed in greater detail by Marsh, Parker, and Barnes (1985). Factor analyses of responses to the SDQ-II by diverse populations of subjects of different ages have each identified the 11 hypothesized factors. To provide a definitive study of this factor structure, a single factor analysis was performed on all 5,494 sets of responses to the SDQ-II in the normative sample. For purposes of this analysis and consistent with most SDQ-II research, the 8 or 10 items from the 11 SDQ-II scales were divided into 4 or 5 item pairs such that the first two items were assigned to the first pair, the next two items to the next pair, and so on. A factor analysis was performed on responses to these 51 item pairs.

Use of Item Pairs

The rationale for using item pairs has been previously discussed by Marsh and O'Neill (1984). Factor analysis of 51 item pairs, instead of the 102 items, is preferable for several reasons. First, a widely used guideline for factor analysis is that the number of subjects should greatly exceed the number of variables. By using item pairs, the ratio of subjects to variables is doubled. Second, the pairing of items to form mini-scales or item-parcels results in more reliable variables that have a smaller component of unique variance than do individual items. Third, factor loadings based on item pairs are less affected by the idiosyncracies of individual items. Fourth, the costs of factor analyses conducted in the development of the SDQ-II and in subsequent use of the instrument are reduced because the number of variables is reduced. And finally, for both the development and later replications of the factor analyses, it is possible to use commercially available computer software (which typically has limits on the number of variables) and hardware (which has limits on memory capacity) to perform the statistical calculations.

Using item pairs also has potential disadvantages: information on individual items is combined and lost; items must be homogeneous (as are the highly intercorrelated SDQ-II items; see Chapter 4); and factor scores may vary somewhat across different factor analyses which have been calculated with different sets of item pairs. However, as Marsh and O'Neill (1984) have shown with the SDQ-III responses, the initial item pairing and a random pairing had factor scores that correlated at a high level (median = .99 across 13 scales). Thus, the item-pairing method appears to be an acceptable practice in analyzing the SDQ-II and its two companion instruments.

Results of the Total Group Factor Analysis

A factor analysis, using iterated communality estimates, a Kaiser normalization, and an oblique rotation to a final solution with delta set to −2.0, was performed on responses to the 51 item pairs derived from the SDQ-II responses. The results of this factor analysis (see Table 5) clearly identify each of the SDQ-II factors. The factor loadings for variables designed to measure each factor, the target loadings, are substantial, ranging from .48 to .80 (median = .68). The nontarget loadings are much smaller, ranging from −.12 to .27 (median = .03). The correlations among the factors are modest, ranging −.03 to .39 (median = .15), and much smaller than the coefficient alpha estimates of reliability (median = .87; see Table 2; p. 25).

The largest correlations occur between the General School and the other academic self-concepts and between the General Self factor and other nonacademic factors. Despite the moderate correlation between General School and Verbal (.41), and between General School and Math (.37), the correlation between Verbal and Math self-concepts (.03) is close to zero. This lack of correlation, though inconsistent with the original Shavelson et al. model, is consistent with other SDQ research that led Marsh and Shavelson (1985) to propose a revision of the original model. Results of factor analyses of responses to the SDQ-II, such as the one presented here, provide strong support for the multidimensionality of self-concept and particularly for the facets hypothesized in the Shavelson et al. model.

Many previous factor analyses of SDQ-II responses have been conducted (e.g., Marsh, 1987c; Marsh, Smith, Marsh, & Owens, 1988) although only the factor analysis by Marsh, Parker, and Barnes (1985) was actually published. Each of these factor analyses was based on responses to the earlier version of the SDQ-II instead of the subset of 102 items that compose the current version. Whereas the 11 SDQ-II factors were well-defined in each of the analyses, there were a few measured variables (i.e., item pairs) with factor loadings below .3. In contrast, the results in Table 5 show that all item pairs had "target" factor loadings (loadings on the scales to which they were assigned) of at least .48. This difference is not surprising in that the purpose of the revision was to reduce the length of the SDQ-II by eliminating the relatively weaker items. The analysis suggests that the revision of the SDQ-II which led to the current set of 102 items has improved the clarity of the factor structure. Table 5 shows the correlations among the 11 factors in the 102-item SDQ-II. Most of the correlations are under .20, which demonstrates the distinctiveness of the SDQ-II factors.

In another analysis the comparability of the 122-item and 102-item versions of the SDQ-II was examined by correlating raw scale scores derived from both versions. That is, for each of the 5,494 subjects in the normative sample, scale scores were computed for the 122 items in the previous version of the SDQ-II and for the subset of 102 items on the current version. Correlations between matching scores based on the old and new versions were at least .985 (median = .994). These results further illustrate that the scores from the two forms are so similar that there is little concern about the generality of findings from previous research to the current revision.

Correlations were also computed between raw scale scores and factor analytically derived scores for matching scales based on responses to the revised SDQ-II. These correlations are all at least .976 (median = .986), suggesting that the factor scores and scale scores are reasonably interchangeable. The main distinction between the raw scale scores and the corresponding factor scores appears to be in the size of the correlations among the scores. The median correlation among the raw scale scores

Table 5. Factor Analysis for the SDQ-II Normative Sample

Variable	Physical Abilities	Physical Appearance	Opposite-Sex Relations	Same-Sex Relations	Parent Relations	Honesty–Trustworthiness	Emotional Stability	Math	Verbal	General School	General Self	h^2
					Factor Loadings							
Physical Abilities 1	**.78**	−.03	.03	.05	.04	.06	.01	.01	.04	−.03	.04	68
Physical Abilities 2	**.71**	.05	.10	.04	.01	−.04	.03	.00	.03	.00	.10	65
Physical Abilities 3	**.75**	.06	.03	.00	.03	.02	.00	.02	−.02	.05	.03	64
Physical Abilities 4	**.67**	.06	.06	−.01	−.01	−.03	.05	.01	.00	.04	.04	53
Physical Appearance 1	.01	**.70**	.18	.09	.02	.01	.01	−.01	.03	.03	.10	74
Physical Appearance 2	.03	**.76**	.16	−.02	.01	−.06	.05	.00	.01	.05	.02	76
Physical Appearance 3	.04	**.74**	.10	.04	.04	.01	.04	.04	.02	.00	.12	77
Physical Appearance 4	.12	**.68**	.08	−.04	.00	−.02	.02	.02	.01	.03	.19	71
Opposite-Sex Relations 1	.04	.09	**.72**	.06	.00	.02	.06	.02	.01	.01	.05	68
Opposite-Sex Relations 2	.08	.24	**.69**	.01	.01	−.05	.01	.00	.00	.04	−.02	71
Opposite-Sex Relations 3	.08	−.04	**.78**	.04	−.01	−.01	.02	.00	.03	−.05	.11	69
Opposite-Sex Relations 4	.04	.20	**.70**	.04	−.01	.00	.01	−.01	−.01	.05	.03	72
Same-Sex Relations 1	.06	.04	.08	**.68**	.02	−.02	.02	.02	.02	.00	.07	58
Same-Sex Relations 2	.02	.17	.08	**.61**	.00	.03	.01	−.02	.05	.11	−.06	51
Same-Sex Relations 3	.03	−.03	.05	**.64**	.04	.02	.03	−.01	.03	.00	.11	51
Same-Sex Relations 4	.01	.02	.12	**.60**	.03	.03	.08	.01	.02	.00	.07	48
Same-Sex Relations 5	.03	−.03	−.02	**.57**	.02	.03	.00	−.02	.05	.06	.10	40
Parent Relations 1	.01	.00	.02	.06	**.77**	.09	.04	.01	.02	.06	.07	73
Parent Relations 2	.03	.03	−.01	−.01	**.77**	−.02	.08	.01	.02	.00	.05	66
Parent Relations 3	.03	.08	−.06	−.02	**.73**	.09	.11	.04	.00	.02	−.02	63
Parent Relations 4	.02	−.04	.04	.03	**.68**	.05	−.08	−.02	.01	.03	.15	56
Honesty–Trustworthiness 1	.00	−.06	.03	.02	.09	**.71**	.02	−.03	.01	.09	.06	61
Honesty–Trustworthiness 2	−.02	.10	−.04	.00	.08	**.65**	.08	.08	.07	.01	−.05	51
Honesty–Trustworthiness 3	.05	−.11	−.06	.02	.10	**.61**	−.12	.08	.07	.06	.10	54
Honesty–Trustworthiness 4	.04	−.09	−.01	.03	.05	**.68**	.00	.05	.08	.05	.06	59
Honesty–Trustworthiness 5	.02	.04	.04	.05	.06	**.61**	.05	.01	.02	.06	.11	52
Emotional Stability 1	.05	.03	.05	.02	.04	.04	**.58**	.01	.01	.04	.09	46
Emotional Stability 2	.03	.00	.05	.05	.04	.01	**.57**	.03	−.01	−.02	.11	47
Emotional Stability 3	.05	.02	.06	.02	−.01	−.05	**.60**	.03	.03	.01	.13	47
Emotional Stability 4	.05	.08	.06	.02	.08	.00	**.66**	.04	.01	.06	.03	58
Emotional Stability 5	.03	.09	.03	.02	.06	.10	**.60**	.02	.02	.07	.02	50

Note. N = 5,494. This table presents the results of a principal-factor analysis with an oblique (direct oblimin) rotation to simple structure. The variables for each factor in the factor loading matrix are the sums of responses to successive item pairs designed to load on that factor. Target loadings are in boldface.

Table 5. Factor Analysis for the SDQ-II Normative Sample *(continued)*

Variable	Factor									
	Physical Abilities	Physical Appearance	Opposite-Sex Relations	Same-Sex Relations	Parent Relations	Honesty–Trustworthiness	Emotional Stability	Math	Verbal	General School	General Self	h^2
Math 1	.03	.02	.02	.00	.00	-.01	.08	**.72**	-.06	.23	-.02	.71
Math 2	.02	.01	.02	-.02	.04	.06	.01	**.77**	.05	.03	.05	.67
Math 3	.02	.05	.00	-.01	.03	.07	.01	**.73**	.00	.14	.02	.68
Math 4	.03	.00	.01	-.01	.03	.05	.00	**.73**	.00	.20	.06	.75
Math 5	.03	.01	-.01	-.02	.01	.04	-.01	**.80**	.02	.08	.07	.75
Verbal 1	.03	.05	.00	.00	.04	.05	-.01	.06	**.75**	-.03	.07	.61
Verbal 2	.02	-.01	.07	.06	.03	-.02	.06	-.09	**.57**	.27	.03	.58
Verbal 3	.01	.03	.01	.05	.00	.05	-.05	.00	**.72**	.08	.02	.62
Verbal 4	.01	-.04	.03	.04	.01	.09	.01	.02	**.65**	.13	.03	.57
Verbal 5	.03	.03	.09	.04	.01	.01	.14	-.06	**.53**	.23	.02	.54
General School 1	.03	.09	.02	.00	.02	.07	.05	.13	.05	**.53**	.09	.52
General School 2	.00	.13	-.04	.06	.02	.05	-.07	.10	.12	**.48**	.11	.47
General School 3	.05	-.01	.04	.00	.02	.09	.05	.09	.11	**.52**	.17	.59
General School 4	.04	.02	-.01	.04	.06	.00	.01	.12	.11	**.64**	.08	.67
General School 5	.05	-.02	.03	.03	.06	.02	.05	.11	.14	**.58**	.16	.67
General Self 1	.05	.15	.06	.04	.08	.04	.05	.02	.01	.05	**.62**	.66
General Self 2	.07	.10	.07	.06	.05	.02	.11	.03	.02	.16	**.53**	.62
General Self 3	.06	.09	.07	.03	.04	.04	.08	.03	.04	.05	**.64**	.66
General Self 4	.12	.09	.08	.10	.12	.02	.11	.00	.03	.09	**.49**	.59
General Self 5	.08	.07	.02	.06	.07	.01	.02	.01	.02	.14	**.57**	.60
Correlations Among Factors												
Physical Abilities	—				—	—	—	—	—	—	—	—
Physical Appearance	.17	—			.25	—	—	—	—	—	—	—
Opposite-Sex Relations	.23	.19	—		.19	.06	.06	.06	—	—	—	—
Same-Sex Relations	.13	.11	.21	—	.09	.16	.06	.03	—	—	—	—
Parent Relations	.11	.08	.04	.10	.10	.18	.07	.03	.41	—	—	—
Honesty–Trustworthiness	.05	-.03	.10	-.02	.17	.21	.15	.37	.18	.39	—	—
Emotional Stability	.16	.18	.12	.19	.29	.18	.27	.14	.18	—	—	—
Math	.08	.08	.01	.00								
Verbal	.08	.07	.08	.16								
General School	.14	.18	.10	.16								
General Self	.28	.31	.26	.25								

Note. N = 5,494. This table presents the results of a principal-factor analysis with an oblique (direct oblimin) rotation to simple structure. The variables for each factor in the factor loading matrix are the sums of responses to successive item pairs designed to load on that factor. Target loadings are in boldface.

is .31, whereas the median correlation among estimated factor scores is .18.[1] These results are similar to those shown in Table 6 where the factor correlations computed directly from the factor analysis have a median of .15. Because one of the aims of the SDQ-II is to maximally differentiate among different components of self-concept, most of the SDQ-II research has used factor scores.

Similarity of the Factor Structures Across Sex and Age

A frequent purpose of factor analytical studies is to demonstrate that factor structures are similar across subgroups. Such demonstrations provide support for the generality of the factors and may be necessary before scores across the subgroups can be justifiably compared. For each of the SDQ instruments, separate factor analyses of responses by boys and by girls has demonstrated the generality of the factor structure across sex. Similarly, separate factor analyses of responses by subjects differing in age have shown the generality of the factor structure across age. (For reviews of this research, see the SDQ-I and SDQ-III manuals; also see Marsh & Hocevar, 1985.)

For purposes of this Manual, the entire set of 5,494 SDQ-II responses in the normative sample was divided into four groups: younger boys, older boys, younger girls, and older girls. Students in grades 7 and 8 were designated as younger; those in grades 9, 10, and 11 were designated as older. Four separate factor analyses—like those conducted for the total group—were conducted for these subgroups. The general pattern of results for each of these four separate factor analyses was remarkably similar to that shown for the total group (see Table 6).

The results of all five factor analyses (the four subgroup analyses and the total group analysis) are summarized in Table 6. In all five factor analyses, the factor structures are very well defined in that target loadings are consistently high, nontarget loadings are consistently low, and factor correlations are moderate. (Target loadings are factor loadings for the variables designed to measure each factor, whereas all other factor loadings are nontarget loadings. For each variable there is only one target loading; that is, each variable is designed to measure only one factor.)

In each of the five factor analyses, the smallest of the 51 target coefficients is never smaller than .42 and is substantially larger than the smallest of the 510 nontarget loadings. Moreover, the nontarget loadings are consistently close to zero (medians vary from .03 to .04). Finally, the moderate correlations among the 11 SDQ-II factors are similar in each of the five factor analyses (medians vary from .14 to .17). Taken together, these findings demonstrate that the SDQ-II factor structures are very similar across age and sex.

To further evaluate the comparability of the factor structures across age-by-sex groups, factor scores derived from the different factor analyses were also compared. For each of the four subgroups, factor scores were derived from the total group factor analysis and from the factor analysis of that subgroup. Thus, two sets of factor scores were derived for each student. One was based on the total group factor analysis and the other on the factor analysis of the subgroup to which the student belonged (e.g., younger boys). Within each of the four subgroups and across all of the subgroups, the median correlation among the 11 matching factor scores was

[1] The finding that factor scores are less correlated than raw scale scores is typical of factor analytic studies and has been observed in research with each of the SDQ instruments (e.g., Marsh, 1986d). It is also worth noting that correlations among the factor scores are slightly higher than the factor pattern correlations in the factor analysis shown in Table 6. This difference is due to technical features of the factor analysis and the estimation of factor scores and is consistent with the results of other factor analytic studies and other SDQ research.

Table 6. Factor Analyses of SDQ-II Responses for the Total Group and Four Subgroups

Variable	Total Group	Boys Grades 7–8	Boys Grades 9–11	Girls Grades 7–8	Girls Grades 9–11
Target Loadings					
No. of Coefficients	51	51	51	51	51
Highest	.80	.77	.78	.79	.84
Lowest	.48	.42	.45	.44	.51
Median	.68	.63	.64	.68	.69
Nontarget Loadings					
No. of Coefficients	510	510	510	510	510
Highest	.27	.26	.29	.35	.26
Lowest	−.12	−.08	−.12	−.13	−.08
Median	.03	.04	.04	.03	.04
Factor Correlations					
No. of Coefficients	55	55	55	55	55
Highest	.39	.43	.41	.41	.40
Lowest	−.03	−.07	.03	−.06	−.04
Median	.15	.17	.15	.16	.14

Note: $N = 5{,}494$.

over .99. These results indicate that these two sets of factor scores were almost perfectly correlated for all SDQ-II factors. The high correlations between factor scores derived from the different factor analyses further support the similarity of the factor structures.

In summary, the results of factor analyses of the four subgroups demonstrate the similarity of the SDQ-II factor structure across sex and age.

Hierarchical Structure of Self-Concept

Whereas the Shavelson et al. model is often depicted as the set of factors used to illustrate the model, Shavelson et al. (1976) actually placed more emphasis on the multifaceted, hierarchical structure than on the number and content of the specific facets. In contrast to a hierarchical structure, Coopersmith (1967) and Marx and Winne (1978) argued that facets of self-concept are so heavily dominated by a general factor that separate components could not be readily differentiated, whereas Soares and Soares (1977) argued that facets are so distinct that no hierarchical ordering existed. Although both of these extremes are actually consistent with a hierarchical model, the model's value is dubious if the hierarchy is so strong that components can barely be distinguished or so weak that components are nearly independent.

In their original model, Shavelson et al. (1976) posited that specific academic self-concepts (e.g., reading, math, science) form a single higher-order facet of academic self-concept. Marsh and Shavelson (1985) tested this model with responses by Australian elementary school children to the SDQ-I but found the hierarchy to be more complicated than anticipated. In particular, Verbal and Math self-concepts were nearly uncorrelated with each other and did not combine with General School self-concept to form a single, second-order academic factor. Instead there were two second-order academic factors representing verbal–academic and math–academic self-concepts that were nearly uncorrelated. This lack of correlation led to a revision of the original model, called the Marsh–Shavelson revision.

Because support for this revision of the Shavelson et al. model was limited to responses by Australians to one SDQ instrument, Marsh, Byrne, and Shavelson (1988) tested its generality with responses to three different instruments by 11th and 12th grade Canadians. The pattern of correlations among the factors was consistent across responses to different self-concept instruments and supported the Marsh–Shavelson revision. In particular, whereas the General School factor was moderately correlated with the Verbal and Math factors, the Verbal and Math factors were nearly uncorrelated with each other. A model developed to account for this lack of correlation between verbal and math self-concepts and their relations to corresponding achievement scores is presented in Chapter 7.

In summary, findings support the contention that self-concept is a multifaceted, hierarchical construct, but they also suggest that the hierarchy may be more complicated and weaker than originally proposed by Shavelson et al.

Summary and Implications

Although many thousands of studies of self-concept have been published, most of the research emphasizes other theoretical constructs, and the interest in self-concept comes from its assumed relevance to these other constructs. These observations help explain why reviews of self-concept research (e.g., Burns, 1979; Shavelson et al., 1976; Wells & Marwell, 1976; Wylie, 1974; 1979) typically emphasize the lack of theoretical basis and the poor quality of measurement instruments.

The multifaceted, hierarchical conceptualization of self-concept apparently is consistent with the perspectives of many researchers, even though it is not reflected in many previously developed instruments. In contrast to the SDQ, most instruments used to measure self-concept have no clearly articulated theoretical basis; this lack makes the examination of construct validity difficult. One approach to this problem has been to test hypotheses from theoretical models (e.g., Marx & Winne, 1978) against responses to existing, largely atheoretical instruments. Because of the generally poor quality of many existing self-concept measures, this approach is dubious. The generally inconsistent results may be due to poor theory, poor test development, or both. The determination of whether or not theoretically consistent and distinguishable facets of self-concept exist should be a prerequisite to the study of how these facets, or overall self-concept, are related to other variables. In adopting this approach, atheoretical or purely empirical approaches to developing and refining self-concept measures were rejected. Instead, an explicit theoretical model was taken as the starting point for scale development, and empirical results were used to support, refute, or revise the instrument *and* the theory on which it is based. For this approach, the Shavelson et al. model was judged to be the best theoretical model of self-concept. Fundamental to this approach is the principle that theory building and instrument construction are inexorably intertwined and that each will suffer if the two are separated. In this sense the SDQ-II is based on a strong empirical foundation and a good theoretical model.

Chapter 6. Age and Sex Effects

Self-concept researchers frequently note that self-concept varies with age, developmental level, and sex (Wylie, 1974, 1979). These differences have diverse theoretical, practical, and methodological implications and have been the focus of many studies. Developmental psychologists are concerned with how self-concept develops and changes during the life span (Harter, 1983) and particularly during the adolescent years when stress is common (Dusek & Flaherty, 1981). Psychologists who focus on sex differences are concerned with how such differences in self-concept develop and change during the life span. The relation between self-concept and sex also appears frequently in psychoanalytic and personality theories (Wylie, 1979). Age and sex effects have important practical implications for the appropriate interpretation of longitudinal studies of self-concept, for the examination of interventions designed to enhance self-concept, and for the construction of normative comparison tables to evaluate self-concept responses. The purposes of this chapter are to examine the effects of age and sex on SDQ-II responses, to compare these results with those based on the SDQ-I (preadolescent) and SDQ-III (late-adolescent) responses, and to relate these findings to other research.

Review of the Literature

There is no consensus in the literature on the effects of sex and age on self-concept, and thus a review of this literature is essential. First, the literature on age effects in self-concept is reviewed. This review is followed by a discussion of sex differences in self-concept.

Age Effects in Self-Concept

Wylie (1979) summarized research conducted prior to 1977 and concluded that the evidence was unconvincing for any age effect, either positive or negative, in overall self-concept in the age range 6 to 50 years. Reports of age effects in specific dimensions of self-concept were too diverse and too infrequent to warrant any generalizations. More recent research, however, suggests a curvilinear effect in which self-concept declines during preadolescence and early adolescence, levels out in middle adolescence, and then increases in late adolescence and early adulthood.

Preadolescence. Soares and Soares (1977) examined self-concepts in children in grades 1 through 8, using two forms of their self-concept instrument that emphasized academic self-concept. For total scores in grades 1 through 3 (Form P) and in grades 4 through 8 (Form I), they found statistically significant declines in self-concept.

Similarly, Trowbridge (1972) reported significant age effects in total self-concept in grades 3 through 8; self-concepts declined in grades 3 through 6 but were stable in

grades 6 through 8. In a cross-sectional study of self-concepts in grades 1 through 4, Eshel and Klein (1981) found that general self-concept scores sharply declined with age.

Other studies have not demonstrated any clear relationship between self-concept and age. Initial research with the *Piers-Harris Children's Self-Concept Scale* (Piers, 1984; Piers & Harris, 1964), for example, indicated that self-concepts in grade 6 were lower than self-concepts in grade 3 (Piers & Harris, 1964) and in grade 4 (Piers, 1965, as cited in Piers, 1984), but this age effect was not replicated in the subsequent research (Piers, 1984). Simmons, Rosenberg, and Rosenberg (1973; Rosenberg, 1985) reported a decline in esteem between the ages of 8 and 13 years. Harter (1982) found no significant age effects in self-concept in grades 3 through 6 for one sample, but results from a larger sample that included students in grades 3 through 9 suggested a decline in self-concept with age. Boersma and Chapman (1979) found no significant age effects in academic self-concept scales administered in grades 2 through 6.

Other researchers have reported significant age effects in self-perceptions of ability in different areas. Nicholls (1979) asked children between the ages of 6 and 12 years to rank their own reading ability compared with that of others in their class and found that these self-rankings declined with age. Stipek (1981) found that children's self-perceptions of their "smartness" dropped between kindergarten and third grade. Ruble, Boggiano, Feldman, and Loebl (1980) reported that self-ratings in a physical ability task (shooting a basketball) were negatively correlated with age in grades 2 through 4.

Marsh, Barnes, Cairns, and Tidman (1984; also see Marsh, 1985a, 1986a) examined age and sex effects for seven self-concept factors measured by the SDQ-I in a cross-sectional study of responses by students in grades 2 through 5. Based on previous research, they hypothesized declines in self-concept with age. Because it is impossible to prove the equivalence of the different age groups in potentially confounding variables, they designed their cross-sectional study so that nonequivalence in age groups worked against their hypothesis of a *linear* age effect. They selected second- and fifth-grade student responses from one set of schools, and third- and fourth-grade responses from another set of schools. The youngest and oldest children in the study came from the same schools; thus, if these students differed systematically from the children from the other set of schools, then the effect would appear to be a *nonlinear* age effect with self-concepts in grades 2 and 5 being systematically higher or lower. For the seven SDQ-I scales and three total scores, the effect of age was statistically significant for all but the Parent Relations scale, and in all but the Peer Relations scale, the significant effects were primarily a linear decline in self-concept with age. This decline was moderate, representing a drop of about one-third of a standard deviation between grades 2 and 5, strikingly linear, and similar for boys and girls. For the Parent Relations scale, there was no age effect, and self-concepts were consistently higher than for the other SDQ-I scales at each age. For Peer Relations there was a significant nonlinear effect in which self-concepts decreased from grade 2 to grade 4, but increased in grade 5. The authors argued that several characteristics of the study made the observed age effects particularly robust. First, the conservative design of the study provided a control against the linear effects of age as a function of nonequivalent age groups. Second, the finding that the highest level of self-concept in grade 2 was reported for Parent Relations, coupled with the lack of age effect for this facet, suggested that the age effects in other areas of self-concept were not an artifact of an age-related response bias.

In summary, the studies considered here suggest a decline in self-concepts during preadolescent years. Although not all of the studies reported such a decline, none reported an increase in self-concept during preadolescence.

Early to Middle Adolescence. Dusek and Flaherty (1981) examined age effects in multiple dimensions of self-concept during adolescence with both longitudinal and cross-sectional samples. For the longitudinal samples, they found little evidence of age effects in self-concept. While age effects in their cross-sectional analyses were significant, these tended to be inconsistent across the three age cohorts that they considered. McCarthy and Hoge (1982) also examined age effects in longitudinal and cross-sectional comparisons of students in grades 7 through 12. For both longitudinal and cross-sectional comparisons, they reported significant increases in self-concept responses. Connell, Stroobant, Sinclair, Connell, and Rogers (1975) examined changes in self-esteem during adolescent years for a large random sample of boys and girls. Boys between the ages of 12 to 18 showed a systematic, primarily linear, improvement in self-concept. For girls, however, the effects were curvilinear. Self-concept initially declined between ages 12 and 13, changed little through about age 17, and then increased. In a review of mathematical constructs, Meece, Parsons, Kaczala, Goff, and Futterman (1982) reported that Math self-concept during junior high and high school years declines steadily, but that the drop for girls begins sooner and is larger than that for boys.

The *Tennessee Self-Concept Scale* (Fitts, 1965; Roid & Fitts, 1988) has been one of the most widely used self-concept instruments in the last 25 years. In the original manual, age effects were not sufficiently well-established to warrant the construction of separate norms. Based on accumulated research since the original publication, Roid and Fitts concluded that there was clear evidence that early adolescents had substantially lower self-concepts than did late adolescents, young adults, and elderly respondents. For this reason, they constructed separate norms for their adolescent sample of junior high school and high school subjects.

As noted earlier, Piers and Harris (1964), Soares and Soares (1977), and Simmons et al. (1973; also see Rosenberg, 1985) all found significant declines in self-concept during preadolescence. Each of these studies also examined self-concepts for older children. Piers and Harris reported a subsequent increase in self-concept in grade 10 compared to grade 6 responses, although no age effects were found in subsequent research (Piers, 1984). Soares and Soares (1977) found no age effects in their grade 9 through 12 sample. Simmons et al. (1973; also see Rosenberg, 1985) reported subsequent increases in self-concept after age 13. Marsh, Parker, and Barnes (1985), in research that is described in greater detail, found that for most of the SDQ-II scales, self-concepts declined in grades 7 through 9, leveled out, and then increased in grades 9 through 11. This U-shaped, quadratic component was statistically significant for 8 of the 11 SDQ-II scales and for the Total Self-Concept score. The nature and even the direction of the age effects, however, differed somewhat, depending on the specific scale. This pattern was replicated in subsequent research reported by Marsh, Smith et al., (1988).

In summary, these studies of self-concept show no consistent pattern of age effects during early and middle adolescence. Some studies suggest curvilinear age effects in which self-concept plateaus at its lowest point during early adolescence, whereas other studies indicate consistent increases or consistent decreases in self-concept during this period.

Late Adolescence and Early Adulthood. Bachman and O'Malley (1977) examined boys' self-esteem in an 8-year longitudinal study based on the Youth and Transition data. Using this large, national probability sample, they assessed self-esteem for a single cohort of boys in 10th, 11th, and 12th grades, and 5 years after most respondents had graduated from high school.

Over this 8-year period self-concepts rose about one standard deviation for boys, with different levels of subsequent educational attainment and different levels of

socioeconomic status. Despite the important strengths of this study, its obvious limitations were its reliance on a single age cohort and the exclusion of female participants.

O'Malley and Bachman (1979), however, noted that similar results were obtained in their National Longitudinal Study of the high school class of 1972 seniors and subsequent follow-ups in 1973, 1974, and 1976, which included responses by both boys and girls. O'Malley and Bachman (1983) also analyzed data from their "Monitoring the Future" project in which large probability samples of high school seniors were tested during the period 1976 through 1979 and followed up either one or two years after initial testing. They again found consistent increases in self-esteem of approximately one-tenth of a standard deviation per year.

In summary, the findings reviewed by O'Malley and Bachman (1983) provide convincing evidence that general self-concept increased steadily during this late-adolescent and early-adult period.

Summary of Age Effects. Despite Wylie's conclusion to the contrary, there appears to be ample evidence of age effects in self-concept responses. The most clearly documented effects are the systematic increases in self-concept during late-adolescence and early-adulthood. The evidence is also good for decreases in self-concept during preadolescence. These results imply a curvilinear age effect in which the decline in self-concept must reverse itself sometime during early or middle adolescence. However, empirical support for this conclusion was mixed. Four studies (Marsh, Parker, & Barnes, 1985; Marsh, Smith et al., 1988; Piers & Harris, 1964; Simmons et al., 1973) reported curvilinear age functions in which self-concept reached its lowest point during this period. Consistent with these conclusions, Rosenberg's (1985) study showed that self-concept disturbances appear to be most acute during *early* adolescence. It should be noted, however, that the studies reviewed here considered only a narrow range of ages within the early-preadolescent to early-adult period, or sampled only a few ages within this range. Furthermore, with the exception of some of the studies reviewed by O'Malley and Bachman (1983), few of the studies considered here met the stringent criteria proposed by Wylie (1979). Finally, because of the emphasis on overall self-concept and the ad hoc nature of specific dimensions that have been considered, the generality of findings based on overall self-concept to more specific facets has not been adequately tested.

Sex Effects in Self-Concept

Wylie (1979), in her comprehensive review of research conducted prior to 1977, concluded that there was no evidence for sex differences in overall self-concept at any age level. She noted, however, that sex differences in specific components of self-concept may be lost when items are summed to obtain a total score. Wylie found, for example, that girls tended to have higher self-reported affiliation than did boys, which was consistent with Maccoby and Jacklin's (1974) study of social self-concept. Meece et al. (1982) also documented that girls, at least by middle-adolescent years, consistently have lower Math self-concepts than do boys.

Other researchers (Dusek & Flaherty, 1981; Marsh, 1989b; Marsh, Smith, & Barnes, 1985) have also suggested that there are counterbalancing sex differences in many specific components—some favoring boys, some favoring girls—that are consistent with traditional sex stereotypes. Dusek and Flaherty, in their longitudinal study of adolescent self-concept, reported sex differences in specific self-concepts that were consistent with sex stereotypes; boys had higher self-concepts in masculinity and achievement/leadership than did girls but lower self-concepts in congeniality/sociability. Harter (1982) found that preadolescent boys had higher physical

self-concepts than did girls but found no sex differences in social, cognitive, or general self-concept scales. In studies of preadolescent responses to an academic self-concept instrument, Boersma and Chapman (1979) found significant differences favoring girls in school satisfaction, reading/spelling, penmanship/neatness, and a total score; however, there were no significant differences for general ability, confidence, and arithmetic. Piers (1984) concluded that there is growing evidence of sex differences in self-concept that are specific in nature. Whereas she found no significant sex differences for total self-concept on the *Piers-Harris Children's Self-Concept Scale*, there were significant sex differences on 33 of 80 items that appeared consistent with sex stereotypes. Meece et al. (1982) suggested that girls have lower Math self-concepts than do boys by junior high and high school years, but they found few reports of sex differences in Math self-concept during primary school years. Stevenson and Newman (1986) found that by tenth grade boys had more positive Math self-concepts and poorer reading self-concepts than did girls, but that sex differences were not statistically significant in grades 1 through 5. Byrne and Shavelson (1986; Marsh, Byrne, & Shavelson, 1988) examined sex differences in responses by students in grades 11 and 12 to various self-concept instruments. Across the three different instruments, boys had higher Math self-concepts, generally higher General self-concepts, lower Verbal self-concepts, and lower academic self-concepts.

Marsh et al. (1984), studying responses to the SDQ-I in grades 2 through 5, found sex differences in Reading and General School favoring girls and in Physical Abilities favoring boys. Marsh, Relich, and Smith (1983), using SDQ-I responses in grades 5 and 6, examined sex differences in coeducational and single-sex schools. In both types of schools, they found that girls had higher self-concepts in Reading and General School, and lower self-concepts in Physical Abilities, Math, and Physical Appearance. For responses by high school students to the SDQ-II, girls tended to have higher scores for the Verbal, Honesty-Trustworthiness, Same-Sex Relations, and, perhaps, General School scales, whereas boys tended to have higher scores in Physical Abilities, Physical Appearance, Math, and, perhaps, General Self and Emotional Stability scales (Marsh, 1987c; Marsh, Parker, & Barnes, 1985; Marsh, Smith et al., 1988). Marsh, Smith et al. (1988) found that a shift from single-sex to coeducational schooling produced an overall increase in self-concept for high school students but had no effect on the relative size of sex differences in specific SDQ-II scales.

In a large random sample of adolescents, Connell et al. (1975) found significant sex differences in responses to the Rosenberg scale favoring boys for all adolescent ages, although the size of such differences was largest during middle-adolescence. O'Malley and Bachman (1979) reviewed and reanalyzed results from several large, nationally representative studies, using variations of the Rosenberg scale. In these studies males consistently had slightly higher esteem (i.e., one-tenth of a standard deviation higher) than did females. These differences were small but reached statistical significance because of the large sample sizes.

Summary. There are small sex effects in favor of males for measures of total self-concept and for measures of esteem derived from the Rosenberg scale. There also appear to be larger, counterbalancing sex differences in more specific facets of self-concept that are generally consistent with sex stereotypes. Sex differences may also depend on age in that some differences (e.g., Math self-concepts favoring boys) appear to be larger in adolescence than in preadolescence. Support for any such age-by-sex interactions, however, is weak.

Age Effects in the Structure and Hierarchy of Self-Concept

In a study particularly relevant to the present investigation, Marsh et al. (1984) examined responses to the SDQ-I for students in grades 2 through 5. Separate factor analyses of responses by students at each of the four grade levels identified the seven factors that the SDQ-I was designed to measure. These findings are generally consistent with other research (Dusek & Flaherty, 1981; Monge, 1973) showing that if the factor structure underlying responses to a particular self-concept instrument is well defined, then the structure is consistent across age groups. Such findings might be used to argue against the Shavelson et al. hypothesis that self-concept becomes more differentiated with age (also see Werner, 1957). Marsh et al. (1984), however, reasoned that for instruments consisting of carefully constructed sets of items specifically designed to measure particular dimensions of self-concept, new, unanticipated dimensions would not likely be identified in responses by older children. Instead, they used the size of correlations among factors at each age level to test the hypothesis of increasing differentiation. They found a consistent decrease in the size of correlations among the factors—particularly for grades 2 through 4—which they interpreted as support for the increasing differentiation of self-concept.

In a reanalysis of these data, Marsh and Hocevar (1985) used structural equation modeling to test the invariance of the factor structure across age and to examine the hierarchical structure of the SDQ responses. They found that, consistent with the original findings, factor loadings were relatively invariant across age, whereas correlations among factors decreased with age. In examining various hierarchical models, they found that neither a single higher-order factor (general self-concept) nor two higher-order factors (academic and nonacademic) were able to explain relations among the seven first-order factors, but that three higher-order factors (verbal/academic, math/academic, and nonacademic) did provide an adequate fit at each of the four grade levels. The inability of the model positing just two higher-order factors to fit the data was due to the surprisingly low correlations between Math and Verbal self-concepts. In subsequent research, near-zero correlations between Math and Verbal self-concepts were found with preadolescent, adolescent, and late-adolescent responses to the SDQ-I, SDQ-II, and SDQ-III (Marsh, 1986d; also see Chapter 7), and with responses to other self-concept instruments (Marsh, Byrne, & Shavelson, 1988). These results are consistent with the Shavelson et al. model in that self-concept was shown to be multifaceted, to be hierarchically ordered, and, perhaps, to become increasingly differentiated with age. The results also suggest, however, that self-concept dimensions are more distinct, that the hierarchy is weaker, and that the hierarchical structure differs from that originally anticipated by Shavelson et al. (see Marsh & Shavelson, 1985).

Age and sex effects in SDQ-II responses were first studied by Marsh, Parker, and Barnes (1985; also see Chapter 7) for a sample of 901 students in grade 7 (mostly 12-year-olds) through grade 12 (mostly 17-year-olds) who attended the same high school in New South Wales, Australia. Traditionally, students leave school after grade 10, at which time a school certificate is awarded. grades 11 and 12 are primarily designed as a two-year program to prepare students to take external examinations near the end of 12th grade. Because there were so few students in grades 11 (N = 48) and 12 (N = 24), Marsh et al. considered them as a single grade level for purposes of their study. They conducted separate analyses of variance (ANOVAs) on each of the SDQ-II scales to determine age and sex effects. The main effects of each variable and a polynomial breakdown of the age effects are summarized in Table 7. A polynomial model divides the age effect into linear ("straight-line"), quadratic ("U-shaped"), and cubic ("S-shaped") trends. Sex-by-age interactions were also considered, but these failed to reach statistical significance for any of the individual scales or Total Self-Concept score.

As shown in Table 7, sex differences were statistically significant for 8 of the 11 scales, but the size and direction of these effects varied with the particular scale. Boys had higher scores for 5 scales (Physical Abilities, Physical Appearance, Emotional Stability, Math, and General Self), whereas girls had higher scores for 3 scales (Same-Sex Relations, Honesty–Trustworthiness, and Verbal). For the remaining 3 scales and for the Total Self-Concept score, the effect of sex was not statistically significant.

Table 7. Sex and Age Effects (%Variance Explained)

	Main Effects		Polynomial Age Trends		
Scale	Sex	Age	Linear	Quadratic	Cubic
Physical Abilities	0.6*	1.6**	n.s.	1.1	n.s.
Physical Appearance	8.9**	1.7**	n.s.	n.s.	1.2**
Opposite-Sex Relations	n.s.	5.6**	5.0**[a]	n.s.	n.s.
Same-Sex Relations	6.8**[b]	1.4*	n.s.	1.9**	n.s.
Parent Relations	n.s.	3.5**	2.1**	1.2**[c]	n.s.
Honesty–Trustworthiness	3.1**[b]	4.7**	n.s.	4.7**[c]	n.s.
Emotional Stability	0.7*	n.s.	n.s.	n.s.	n.s.
Math	2.2**	1.8**	n.s.	1.3**[c]	n.s.
Verbal	3.5**[b]	1.4*	n.s.	0.8*[c]	0.6*
General School	n.s.	1.1*	n.s.	0.7*[c]	n.s.
General Self	2.1**	1.6**	n.s.	1.0**[c]	n.s.
Total Academic	n.s.	2.0**	n.s.	1.6*[c]	n.s.
Total Self-Concept	n.s.	3.1**	0.6*[a]	2.8**[c]	n.s.

Note. N = 901.

[a]Denotes age effects in which the linear component showed an increase in self-concept with age; in all other cases self-concept declined with age or was nonsignificant in either direction.

[b]Denotes sex differences in which females had higher self-concepts; in all other cases males had higher self-concepts.

[c]Denotes age effects in which the quadratic component indicates a "U-shaped" relationship where self-concepts first decreased with age and then increased.

*$p < .05$; **$p < .01$.

From "Multidimensional Adolescent Self-Concepts: Their Relationship to Age, Sex and Academic Measures," by H. W. Marsh, J. Parker, & J. Barnes, 1985, *American Educational Research Journal, 22*, p. 435. Copyright 1985 by the American Educational Research Association. Adapted by permission of the publisher.

Age effects were significant for 10 of the 11 scales. The linear effects of age were significant for only two scales—Opposite-Sex Relations improved with age, whereas Parent Relations declined. For most of the scales and the Total Self-Concept score, the primary effects of age were in the quadratic component. Self-concepts were highest in grades 7, 11, and 12 and were typically lowest in grade 9. It is important to note that self-concept had begun to improve by grade 10 so that the unrepresentative sample of students in grades 11 and 12 was not responsible for the quadratic effects.

Sex and Age Effects in SDQ-I, SDQ-II, and SDQ-III Responses

The purpose of the analyses described in this section is to examine age and sex effects in the SDQ-II responses and to relate these to effects observed with SDQ-I and SDQ-III. In the first set of analyses, the effects of age and sex on responses to each instrument are examined. Second, the relation between age and the differentiation among the scales is examined.

The first set of analyses is based on responses to the SDQ-I (N = 4,362 sets of responses by 3,679 students in grades 2 through 9)[1], the SDQ-II (N = 5,494 sets of responses by 3,073 students in grades 7 through 11), and the SDQ-III (N = 2,410 sets of responses by 1,202 individuals aged 15 and older). These numbers represent the normative samples for each of these instruments. The samples and published studies based on each instrument are described in the respective SDQ manuals. The normative data bases for both the SDQ-I and SDQ-II, though not collected according to a specific sampling design, appear to be broadly representative of school-aged children in greater metropolitan Sydney, Australia. The representativeness of the normative data for the SDQ-III is more dubious in that it is based largely on responses by Outward Bound participants (Marsh, Richards, & Barnes, 1986a, 1986b), grades 11 and 12 students in a Catholic girls' school (Marsh & O'Niell, 1984), and university students (Marsh, Barnes, & Hocevar, 1985). It is important to emphasize that the SDQ-I, SDQ-II, and SDQ-III differ in the number of response categories allowed (five, six, and eight respectively), in the scales that are included, and in the specific wording of items. For this reason, responses from the three instruments cannot easily be incorporated into a single analysis. For the present purposes, separate analyses were conducted on the responses to each SDQ instrument, and the pattern of results was then compared across the different analyses. Because subjects in some studies completed SDQ instruments on more than one occasion, the number of completed instruments is substantially larger than the total number of respondents. This issue was addressed by assigning each individual a weight equal to one over the number of times the individual completed an instrument, and by conducting subsequent analyses on the weighted responses.

Main and Interaction Effects

Separate sets of ANOVAs were conducted on responses to each of the SDQ instruments to determine the effects of sex and age. For the SDQ-I and SDQ-II responses, age was taken to be grade in school, whereas age for the SDQ-III responses was divided into three discrete categories (younger than 18 years, 18 to 21.5 years, and older than 21.5 years). In each set of the analyses, the main effect of sex is a single degree-of-freedom contrast. The main effect of age is summarized by three single degree-of-freedom contrasts consisting of the linear, quadratic, and cubic effects of age for the SDQ-I and SDQ-II responses, and consisting of the linear and quadratic components for the SDQ-III responses. The sex-by-age interaction is summarized by the interaction of sex and the single degree-of-freedom age contrasts for the linear, quadratic, and cubic trends. Because of the large number of tests and large sample sizes, statistical significance was set at the .01 level, but the variance explained by all effects is summarized in Table 8. Separate analyses were conducted for both raw scale and factor scores, but the results were so similar that only the results of the raw score analyses are reported in Table 8.

[1] Note that the sample size (N = 4362) cited here differs substantially from the normative data sample reported in the SDQ-I Manual. The present sample includes all of the norm sample as well as a substantial number of additional responses by students in grades 7, 8, and 9 collected subsequently to the development of the SDQ-I norms.

Age Effects. For the range of preadolescent ages of subjects responding to the SDQ-I, there is a clear linear decline in self-concept with age. For all SDQ-I scales and Total Self-Concept score, this decline is statistically significant, primarily linear, and occurs for both boys and girls (see Figure 14).

For the range of early and middle adolescent ages of subjects responding to the SDQ-II, there is a reasonably consistent quadratic (U-shaped) effect; self-concepts are relatively higher in grade 7, decline in grades 8 and 9, and then increase in grades 10 and 11. This quadratic effect is statistically significant for 8 of the 11 SDQ-II scales and the Total Self-Concept score and occurs for both boys and girls. In contrast to the consistency of the quadratic effects, the direction of the linear age effects on SDQ-II responses is not consistent across different scales. The linear effects are positive for some scales and negative for others, and there is no linear effect for the Total Self-Concept score. The variance explained by age in the SDQ-II data tends to be much smaller than was the case with the SDQ-I or SDQ-III data.

For the range of late-adolescent and early-adult subjects responding to the SDQ-III, the increase in self-concepts with age is reasonably consistent. The linear age effect is statistically significant for 11 of the 13 SDQ-III scales, and the direction of this effect is positive for 9 of these scales and for the Total Self-Concept score.[2]

In summary, particularly for the Total Self-Concept scores, there is a consistent pattern of age effects across the three SDQ instruments. During preadolescence (SDQ-I data), there is a linear decline in self-concept. During early and middle adolescence (SDQ-II data), self-concept continues to decline through about grade 8 or 9, levels out, and then increases in grades 10 and 11. During late adolescence and early adulthood (SDQ-III data), self-concept continues to increase. This overall trend occurs for both males and females, although its nature varies somewhat for specific scales.

Sex Effects. For most of the SDQ-I, SDQ-II, and SDQ-III scales, sex effects are statistically significant; some favor girls, but more favor boys. Reflecting this tendency, the Total Self-Concept score favors boys, although this sex difference consistently explains only about 1 percent of the variance in each of the three data sets. The direction of sex differences in specific scales tends to be consistent with traditional sex stereotypes.

For the six scales common to the three instruments (Physical Abilities, Physical Appearance, Parent Relations, Math, Verbal, and General School), stereotypic sex differences are reasonably consistent across the three data sets: (a) boys have higher Physical Abilities, Physical Appearance, and Math self-concepts for all three data sets; (b) there are no sex differences for the Parent Relations scale in any of the three data sets; and (c) girls tend to have higher Verbal (for SDQ-I and SDQ-II data) and General School (for SDQ-I and SDQ-III data) self-concepts.

[2] The negative effect of age on Religion and the age-by-sex interaction for the SDQ-III data should be interpreted cautiously because a majority of the female respondents in the youngest SDQ-III age group were students in a Catholic girls' high school. Thus, the higher Religion self-concept scores for the youngest age group are likely an artifact of sampling.

Table 8. Summary of Sex and Age Effects (%Variance Explained) for SDQ-I, -II, and -III on Raw Scale Scores

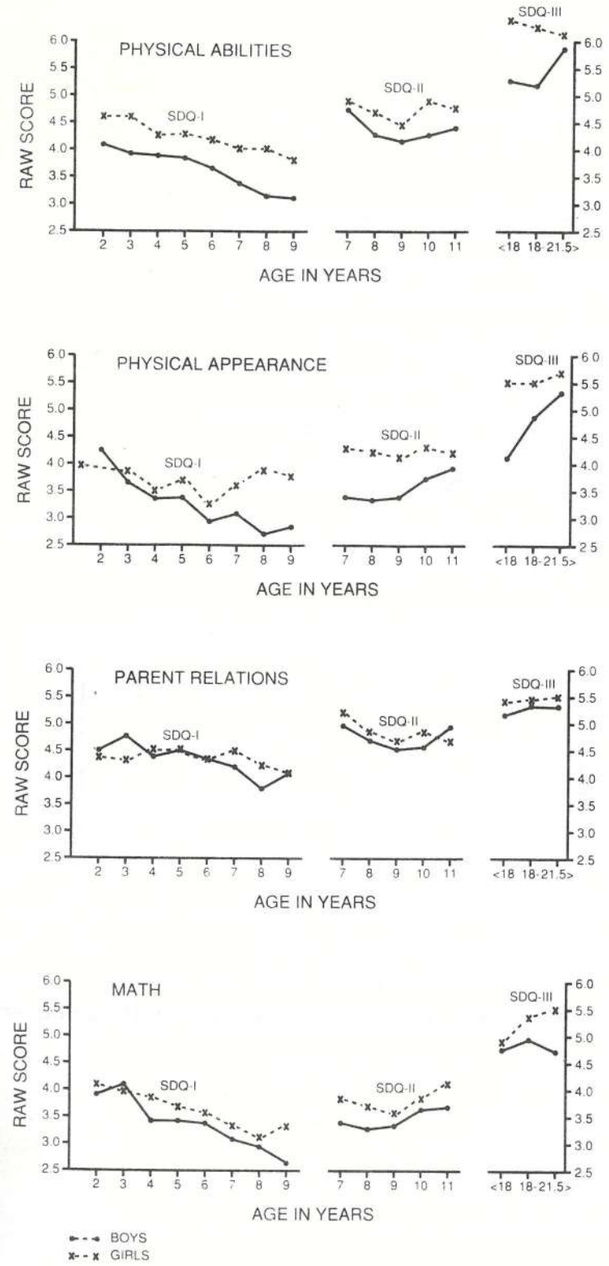
	Main Effects	Polynomial Age Trends			Interaction Effects		
Scale	Sex	Linear	Quadratic	Cubic	Sex × Linear Age	Sex × Quadratic Age	Sex × Cubic Age
SDQ-I							
Physical Abilities	8.89**	4.66**	0.12	0.01	0.17*	0.17*	0.01
Physical Appearance	2.63**	3.00**	1.14**	0.25*	1.40**	0.03	0.01
Peer Relations	0.83**	1.40**	0.14	0.00	0.05	0.00	0.00
Parent Relations	0.03	2.97**	0.63**[a]	0.02	0.04	0.00	0.01
Reading	1.40**[b]	3.30**	0.04	0.01	0.12	0.03	0.00
Mathematics	0.99**	3.99**	0.00	0.00	0.03	0.00	0.06
General School	0.09[b]	6.70**	0.04	0.01	0.13	0.05	0.03
Total SDQ-I	0.92**	8.61**	0.05	0.01	0.40**	0.05	0.01
SDQ-II							
Physical Abilities	2.16**	0.58**	0.89**	0.13	0.06	0.03	0.01
Physical Appearance	7.81**	0.30*[c]	0.19*	0.10	0.17	0.05	0.00
Opposite-Sex Relations	1.15**	1.83**	0.00	0.05	0.26*[d]	0.13	0.03
Same-Sex Relations	2.76**[b]	0.29**	0.62**	0.73**	0.16	0.33*	0.02
Parent Relations	0.20[b]	1.28**	0.67**	0.04	0.18	0.10	0.06
Honesty–Trustworthiness	3.61**	0.06	1.45**	0.04	0.17	0.01	0.00
Emotional Stability	2.35**	0.00	0.04	0.00	0.00	0.07	0.38*
Math	1.58**	0.13	0.54**	0.05	0.01	0.01	0.05
Verbal	0.77**	0.00	0.04	0.15	0.06	0.14	0.26*
General School	0.09	0.29*	0.42**	0.33*	0.06	0.01	0.02
General Self	1.05**	0.12	0.59**	0.26*	0.19	0.05	0.02
Total SDQ-II	1.04**[b]	0.00	1.00**	0.31*	0.19	0.00	0.07
SDQ-III							
Physical Abilities	4.90**	1.55**[c]	0.02	—	0.83**[d]	0.14	—
Physical Appearance	7.64**	12.0**[c]	1.35**[a]	—	2.12**[d]	0.24	—
Opposite-Sex Relations	0.20[b]	0.16	0.15	—	1.32**[d]	0.13	—
Same-Sex Relations	0.01	0.90**	0.00	—	0.22	0.03	—
Honesty–Trustworthiness	0.44*[b]	2.14**[c]	0.02	—	0.07	0.20	—
Parent Relations	0.12	0.17	0.05	—	0.00	0.02	—
Spiritual Values–Religion	4.42**[b]	1.14**	0.25	—	1.30**	0.16	—
Emotional Stability	2.04**	2.85**[c]	0.01	—	0.50	0.30	—
General Self	2.35**	2.11**[c]	0.21	—	0.45	0.13	—
Verbal	0.01	1.77**[c]	0.00	—	0.00	0.79*	—
Mathematics	2.21**	1.61**[c]	0.07	—	0.57*	0.00	—
Academic	0.02[b]	2.98**[c]	0.00	—	0.00	0.28	—
Problem Solving	5.12**	3.23**[c]	0.27	—	0.04	0.06	—
Total SDQ-III	1.07**	3.62**[c]	0.10	—	0.13	0.14	—

Note: $N = 4{,}362$ (SDQ-I); $N = 5{,}494$ (SDQ-II); $N = 2{,}410$ (SDQ-III). A series of two-way analyses of variance was conducted in which the separate contrasts were used to test the effects of sex; the linear, quadratic, and cubic components of age; and the sex interaction. Effect sizes, the percentage of variance explained (i.e., eta squared × 100%), are all based on single degree-of-freedom contrasts. Graphs of scores common to the three instruments (Physical Abilities, Physical Appearance, Parent Relations, Verbal/Reading, Math, School/Academic, and Total Self-Concept) appear in Figure 14.

[a]The quadratic effects of age are negative (i.e., the slope becomes more negative or less positive with age as in an inverted "U-shaped" effect); other significant quadratic age effects are positive (i.e., the slope becomes more positive or less negative with age as in a "U-shaped" effect).
[b]Girls have significantly higher self-concepts than do boys; other significant sex effects are in favor of boys.
[c]The linear effect of age is positive; other significant linear age effects are negative.
[d]The significant linear-age × sex interactions indicate that sex differences shift in the favor of girls with age (i.e., larger differences in favor of girls or smaller differences in favor of boys); other significant linear-age × sex interactions indicate a shift in favor of boys with age.

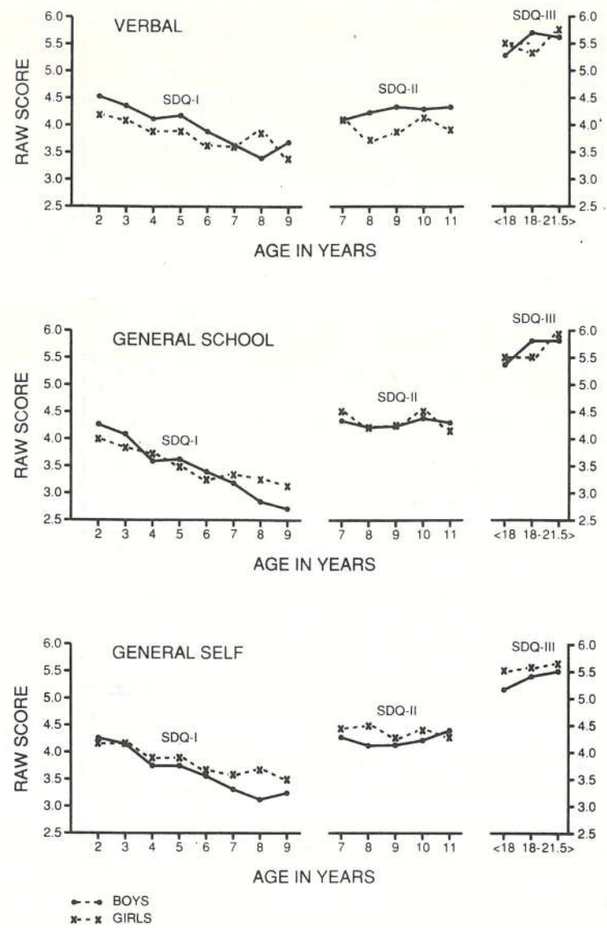
*$p < .01$, **$p < .001$.

From "Age and Sex Effects in Multiple Dimensions of Self-Concept: Preadolescence to Early-Adulthood" by H. W. Marsh, 1989, *Journal of Educational Psychology, 81*, p. 423. Copyright 1989 by the American Psychological Association. Adapted by permission of the publisher.



Note. The response ranges for the instruments differ: SDQ-I = 5; SDQ-II = 6; SDQ-III = 8.

Figure 14. Age and Sex Effects on the Scales Common to the Three SDQ Instruments



Note. The response ranges for the instruments differ: SDQ-I = 5; SDQ-II = 6; SDQ-III = 8.

Figure 14. Age and Sex Effects on the Scales Common to the Three SDQ Instruments

For the scales not common to the three instruments, differences also appear to be consistent with sex stereotypes. For the SDQ-II and SDQ-III data, boys have higher Emotional Stability, Problem Solving, and General Self scores, whereas girls have higher Honesty–Trustworthiness and Religion–Spiritual Value scores. Sex differences on the social scales, however, are mixed and not fully consistent with traditional sex stereotypes favoring girls. Boys have somewhat higher Peer Relations scores on the SDQ-I and Opposite-Sex Relations scores on the SDQ-II (although this difference occurs only for the younger students in the SDQ-II data). Girls have higher Same-Sex Relations scores on the SDQ-II and slightly higher Opposite-Sex Relations scores on the SDQ-III (although this difference occurs only for the oldest two groups in the SDQ-III data). Thus, even for the social scales, there may be a trend favoring girls that would be consistent with traditional sex stereotypes.

It was anticipated that sex differences might vary substantially with age. In these three data sets, however, few age-by-sex interactions are statistically significant (see Table 8), and those that are typically account for less than 1 percent of the variance. In fact, across all three data sets, Physical Appearance is the only scale in which sex differences vary substantially with age. These findings suggest that sex stereotypes have already affected self-concepts by preadolescence and that these effects are relatively stable from preadolescence to early adulthood.

Age Effects on Differentiation Among Self-Concept Facets

Shavelson et al. (1976) posited that self-concept becomes more differentiated with age. Marsh et al. (1984) found support for this proposal in that the average correlation among SDQ-I scales decreased dramatically with age during the early preadolescent ages (grades 2 through 5). The purpose of analyses to be considered here is to test the generality of these findings across the preadolescent to early adult period and to consider alternative tests of the Shavelson et al. prediction.

Size of Correlations Among Different Scales. Marsh et al. (1984) proposed that correlations among SDQ scales should become smaller with age. This hypothesis was examined by determining at each age level the mean correlation among (1) all scales measured by each SDQ instrument; (2) the six scales common to the three SDQ instruments; and (3) seven correlations chosen a priori on the basis of theory and previous research to be the smallest (Physical Abilities with Verbal, Physical Abilities with Math, Physical Abilities with School, Physical Appearance with Verbal, Physical Appearance with Math, Physical Appearance with School, and Math with Verbal; see earlier discussion) among those common to the three SDQ instruments.

Correlations based on both raw scale scores and factor analytically derived scores were examined. The pattern of correlations at each age is consistent across the three sets of correlations and across the scale and factor scores (see Table 9). At each age level, (a) the mean correlation among raw scores is higher than the mean correlation among factor scores; (b) the mean correlation among all scales is similar to the mean correlation among just the scales common to the three instruments; (c) the mean correlation among scales selected a priori to be lowest is substantially lower than the mean correlation among all scales or among scales common to the three SDQ instruments. This consistency facilitates the comparison of correlations across the different ages.

Consistent with the findings by Marsh et al. (1984), the size of correlations substantially decreases from grade 2 to grade 3 and decreases less between grades 3 and 4 and between grades 4 and 5. There is no support, however, for any further declines in the average correlation among scales for the rest of the SDQ-I data or for

Table 9. Differentiation of Self-Concept with Age: SDQ-I, -II, and -III

Age Level	Sample Sizes[a]		Mean Correlations						Standard Deviations			
			All Scales		Common Scales		Selected Scales		All Scales		Common Scales	
	Weighted	Unweighted	Raw	Factor	Raw	Factor	Raw	Factor	Raw	Factor	Raw	Factor
SDQ-I												
Grade 2	176	176	.55	.43	.55	.43	.49	.37	.52	6.50	.52	6.36
Grade 3	107	107	.37	.27	.39	.29	.30	.20	.62	6.91	.61	6.59
Grade 4	360	513	.34	.24	.31	.20	.23	.12	.75	8.07	.78	8.09
Grade 5	1428	1697	.27	.18	.25	.15	.18	.08	.75	8.02	.78	8.07
Grade 6	1111	1378	.28	.18	.25	.16	.17	.07	.81	8.42	.84	8.53
Grade 7	151	151	.39	.28	.38	.27	.31	.19	.74	7.70	.76	7.76
Grade 8	161	161	.33	.22	.31	.21	.29	.17	.80	8.82	.83	8.93
Grade 9	179	179	.28	.19	.27	.18	.23	.13	.78	8.62	.81	8.89
Total	3679	4362	.33	.22	.31	.20	.24	.13	.76	8.11	.79	8.17
SDQ-II												
Grade 7	929	1374	.30	.18	.29	.17	.21	.09	.98	8.55	1.04	8.28
Grade 8	630	1323	.32	.19	.31	.19	.26	.12	.96	8.49	1.03	8.29
Grade 9	528	1116	.32	.19	.29	.16	.24	.11	.92	8.39	.99	8.38
Grade 10	705	1187	.31	.18	.28	.16	.21	.09	.90	8.05	.96	8.10
Grade 11	281	494	.28	.16	.24	.12	.19	.07	.87	7.99	.95	8.18
Total	3073	5494	.31	.18	.29	.17	.23	.10	.94	8.34	1.00	8.25
SDQ-III												
<18 yrs	377	531	.21	.10	.20	.10	.10	.00	1.16	8.93	1.17	9.12
18–21.5 yrs	407	919	.28	.16	.27	.14	.21	.08	1.07	8.18	1.00	7.96
>21.5 yrs	418	960	.27	.16	.26	.13	.22	.10	1.07	8.04	.96	7.65
Total	1202	2410	.25	.14	.26	.14	.20	.08	1.10	8.37	1.04	8.22

Note: The mean correlation was computed separately for each of three sets of scale scores: All = all correlations among the scales measured by a particular instrument; common = all correlations among the set of six scales common to all three instruments; and selected = correlations selected a priori to be the lowest. The standard deviation of responses by each respondent was also computed for all scales and for the set of common scales (i.e., a respondent who had the same score for all scales would have an *SD* of 0). Separate analyses were conducted for raw scale scores and factor scores.

[a]Unweighted sample sizes refer to the number of instruments completed, and weighted sample sizes refer to the number of different respondents (i.e., many respondents completed an instrument more than once). Analyses presented here are based on the weighted samples (i.e., the weight assigned to each respondent was 1 divided by the number of instruments completed by that respondent).

Adapted from "Age and Sex Effects in Multiple Dimensions of Self-Concept: Preadolescence to Early-Adulthood," by H. W. Marsh, 1989, *Journal of Educational Psychology, 81*, p. 426. Copyright 1989 by the American Psychological Association. Adapted by permission of the publisher.

the SDQ-II and the SDQ-III data. Across the various comparisons, the mean correlation among SDQ scales is as high or higher in grades 6, 7, 8, and 9 as it is in grade 5. For the SDQ-II data, the size of correlations based on all the comparisons is reasonably similar for different ages (there is a weak tendency for correlations to be smaller for the youngest and the oldest subjects completing the SDQ-II). For the SDQ-III data, correlations among scales are consistently smaller for the youngest age group. Thus, the data do not support the Shavelson et al. prediction that self-concept becomes more differentiated with age beyond early preadolescence.

Within-Subject Standard Deviations. Correlations provide a measure of the relative agreement among scores but are insensitive to mean differences among scores. An alternative test of the Shavelson et al. prediction is that differences among the scale scores for the same person will become larger with age. That is, younger subjects are more likely to have uniformly high or uniformly low self-concepts across all areas, whereas older subjects are more likely to have relatively high self-concepts in some areas and relatively low self-concepts in other areas. This hypothesis was tested by computing the standard deviation of scale scores for each subject (i.e., a within-subject standard deviation) for all scales and for the scales common to the three SDQ instruments. Separate within-subject standard deviations were computed for raw scale scores and for factor scores (that were standardized to have a mean of 50 and a standard deviation of 10 across all respondents to each instrument).

For the SDQ-I data, within-subject standard deviations increased between grades 2 and 3 and between grades 3 and 4, but they did not increase systematically for older subjects. For both the SDQ-II and SDQ-III data, within-subject standard deviations decreased with age instead of increasing with age as predicted. Although caution must be exercised in comparing within-subject standard deviations at different ages, the present comparisons offer no support for the Shavelson et al. prediction beyond early preadolescence.

Summary and Implications

The purposes of the present investigations were to examine (a) sex and age effects in SDQ-II responses as compared with those in SDQ-I and SDQ-III responses and (b) alternative methods for operationalizing the Shavelson et al. prediction that self-concept becomes more differentiated with age. Despite claims that self-concept does not vary with age (e.g., Wylie, 1979), the self-concept literature provides strong support for the increase of self-concept during late adolescence and reasonably good support for its decline during preadolescence. This pattern logically necessitates a quadratic, U-shaped effect during adolescence, but there is only weak support for this effect. The problem appears to be that most researchers have considered only a limited age range, or only a few different ages across a wide age range, as well as the many other methodological problems that have plagued self-concept research. The results of the studies just summarized are important because they provide clear support for the predicted age effects on self-concept that were pieced together from a collage of different studies. The results suggest that this effect is reasonably consistent across well-differentiated self-concept scales and across responses by males and by females, based on responses to well-standardized, multidimensional instruments.

For the Total Self-Concept score, the results of the present investigation suggest that males have somewhat higher self-concepts than do females. Marsh and Shavelson (1985) argued, however, that self-concept cannot be adequately understood if only a global component is considered. Consistent with this claim, studies

by Wylie (1979) and others have suggested that the relatively weak sex effects in global self-concept may be due to a counterbalancing of sex differences in more specific areas, some favoring boys and some favoring girls. Because self-concept researchers have typically considered only a global self-concept, or have measured multiple dimensions of self with ad hoc instruments, there has not been a good empirical basis for testing this suggestion. Common sense and preliminary empirical research suggest that sex differences in specific areas of self-concept are consistent with traditional sex stereotypes. The results of the present investigation provide strong support for this contention and demonstrate that the posited effects are reasonably consistent across the preadolescent to early adult period considered here.

The lack of support for the increased differentiation of self-concept with age beyond that previously found during early preadolescent years (Marsh et al., 1984; Marsh & Hocevar, 1985) was surprising. The limited support that was found among preadolescents and the lack of support for older subjects require further consideration. Although Shavelson et al. (1976) predicted increased differentiation with age, they did not devise a clear test of this proposal. Some researchers have argued that it is difficult to test this hypothesis with fixed-format items like those used on the SDQ instruments (e.g., Montemayor & Eisen, 1977), whereas others have used the comparability of factor structures across age to test the hypothesis. These tests of the differentiation hypothesis are new, and, thus, they may require closer scrutiny. Although the hypothesis was supported by the findings for early preadolescent children, this support also warrants further consideration. Because responses by these young children are so very skewed and because some of these young children may have difficulty responding appropriately on a 5-point scale (see, however, Marsh, 1988, in press), other explanations of the findings are plausible. Thus, it is premature to conclude either that there is increased differentiation in self-concept during the early preadolescent years or that there is no increased differentiation beyond early preadolescence.

What are the implications of the decline in self-concepts that occurs in preadolescence and early adolescence? Although it is tempting to seek explanations for this decline in self-concept and to look for culprits (such as schools, parents, society), it should not be seen as "bad" or unfortunate. Indeed, the very high self-concepts of the youngest children are apparently unrealistically high, and, perhaps, it would be unfortunate if their self-concepts did not become more realistic on the basis of additional life experience. For example, Stipek (1981; Stipek & Tannatt, 1984) reports that in interviews with 96 children at the start of first grade, all students claimed to be among the smartest in their class. Even if the self-concepts of the youngest children are unrealistic, however, this does not mean that their self-concepts, or responses to the SDQ, are biased. To the contrary, so long as their responses accurately reflect their self-perceptions, whether or not these self-perceptions are realistic when judged by external standards, the interpretations based on their self-concept responses are valid. Instead, the bias lies in the inferred self-concepts based on the observations of external observers or other indicators that do not reflect this age effect. Future research is needed to identify what characteristics validly affect self-concept, to develop theoretical perspectives consistent with these effects, and to explore the implications of these theoretical and empirical findings.

In one such developmental proposal (also see Harter, 1983; Nicholls, 1979; Ruble et al., 1980; Stipek, 1981, 1984; Stipek & Tannatt, 1984; Werner, 1957, for related theoretical positions), Marsh (1985a; 1989a) posited that very young children are egocentric and have consistently high, less-differentiated self-concepts in all areas. As children become older, they incorporate more external information into their self-concepts so that their self-concepts become more correlated with external criteria. This finding implies that for most children, self-concepts will generally

decline with age; moreover, self-concepts will also become more differentiated so that certain facets are higher while others are more moderate in keeping with the child's external experiences. Marsh (1986d) also demonstrated that children form self-concepts in specific areas by comparing their own abilities in different areas as well as comparing their abilities with those of others. Thus, for example, children who are well above average in all school subjects may have an average self-concept in the subject in which they perform least well. As children incorporate more information and feedback about their actual skills and abilities into the formation of their self-concepts in different areas, their self-concepts will also become more differentiated. Thus, this proposal is consistent with (a) the decline in preadolescent self-concepts with age; (b) the increased differentiation of self-concepts with age at least during early preadolescence; (c) the finding that as children become older, their self-perceptions become more highly correlated with performance, performance feedback, and other external criteria; and (d) the finding that specific dimensions of academic self-concept (e.g., Verbal/Reading and Math) are much more distinct than corresponding achievement in these academic subjects.

Age and sex effects observed here were generally consistent with predictions, but it was surprising to find that sex differences did not vary more with age. This pattern suggests, perhaps, that traditional sex stereotypes have already affected the self-concepts by early preadolescence and that these effects remain relatively stable through at least early adulthood. Self-concept associated with Physical Appearance, however, was a notable exception to this pattern. Whereas the youngest girls (grade 2) had higher Physical Appearance self-concepts than did the youngest boys, the decline in the girls' scores was particularly dramatic so that by the end of preadolescence, girls had substantially lower scores than did boys. This substantial sex difference remained stable across adolescent years but grew smaller during late adolescence and early adulthood. Although other researchers have speculated about such differences, some instruments confound physical attractiveness and physical/athletic ability. Recent research (Marsh, 1987b; Marsh & Jackson, 1986) has shown that these components are quite distinct, particularly for girls, and should not be incorporated into a single physical scale. In the present investigation, even the youngest girls had substantially lower Physical Abilities self-concepts than did boys, whereas during adolescence, sex differences favoring boys are much larger for Physical Appearance than for Physical Abilities. The intriguing question requiring further research is why very young girls think of themselves as more attractive than do boys, whereas girls in the middle preadolescent to early adult period think of themselves as much less attractive than do boys. Perhaps, for this one area, the effect of sex stereotypes may vary with age or the nature of sex stereotypes may differ with age (e.g., all little girls are pretty, but starting at an early age girls are encouraged by society to compare themselves with unrealistic standards of physical attractiveness).

The focus of these analyses has been on theoretical issues related to the effects of age and sex on SDQ-II responses as well as on SDQ-I and SDQ-III responses. The results also have practical implications for the interpretation of SDQ-II responses. Because sex effects for some of SDQ-II scales are substantial, separate norms tables were constructed for responses for males and females. Because age effects and their interaction with sex effects are typically small, separate norms tables for adolescents at different ages were not deemed necessary.

Chapter 9. Relationship of Self-Concept to Masculinity,Femininity, and Androgyny

Several studies have related global measures of self-concept to measures of masculinity, femininity, and androgyny (individuals who are high on both masculinity and femininity) in order to test theoretical predictions of androgyny theory. Unfortunately, these studies typically have not incorporated the recent emphasis on multidimensionality in self-concept. Marsh (1987c) reviewed theories of relations between self-concept responses and measures of masculinity (M) and femininity (F) and described a new model that did emphasize the multidimensionality of self-concept. In contrast to existing theories, this new model posited that the relative contribution of M and F to different dimensions of self-concept would vary systematically with the particular dimension of self-concept. This new model was tested with responses to SDQ-II. The purpose of this chapter is to summarize the results of this study.

The Androgyny Construct

Virtually all researchers prior to 1973 and many personality inventories still in use today assume that M and F are the endpoints of a single, bipolar dimension. More recently Constantinople (1973), Bem (1974), Spence (1984), and others have argued that it is logically possible to possess both M and F characteristics. The existence of both of these traits in the same person has been labeled *androgyny*. The two key assumptions to Bem's 1974 theoretical description of androgyny are that M and F are distinguishable, orthogonal dimensions, and that individuals high on both are mentally healthier and socially more effective.

The *Bem Sex Role Inventory* (BSRI; Bem, 1974) and the *Personal Attributes Questionnaire* (PAQ; see Spence, 1984) are the most widely used androgyny instruments, but their reliance on only socially desirable attributes may constitute an important weakness (see Marsh, 1987c). In response to this potential weakness, Spence, Helmreich, and Holahan (1979) developed an expanded PAQ (EPAQ) to include socially *undesirable* characteristics, and Antill, Cunningham, Russell, and Thompson (1981) developed the *Australian Sex-Role Scale* (ASRS) to measure M and F with positively valued characteristics (M+ and F+) and with negatively valued characteristics (M− and F−). The ASRS was used in the present investigation.

In support of androgyny theory, researchers have typically found that M/F correlations differ significantly from a perfect negative correlation of −1.0. Marsh and Myers (1986; Marsh, 1985b), however, found that the correlations between M and F varied from moderately positive to close to −1.0, depending on the instrument used to assess M and F. Other investigators have found that responses to the adjectives "masculine" and "feminine" are substantially negatively correlated and form a two-item bipolar factor (e.g., Pedhazur & Tetenbaum, 1979). Hence, empirical M/F correlations apparently will not provide unequivocal support for either the androgyny or the bipolar assumptions. Support for androgyny theory requires the

demonstration that both M and F contribute uniquely and positively to the prediction of an appropriate criterion. The most frequently studied criterion has been global measures of self-esteem.

Theoretical Models

Five different models of Masculinity–Femininity and self-concept relations were examined. These models can be tested by assessing the main and interaction effects of M, F, and gender.

Sex-Typed Model. This model posits that the acquisition of a masculine identity by males and a feminine identity by females leads to higher self-concepts and that support for the model would be strongest during early adolescent years when such acquisition processes are presumed to be most important (e.g., Kohlberg, 1966; Lamke, 1982). Support for this model requires that the effects of M and F each interact with gender in an appropriate manner.

Additive Androgyny Model. This model, as described by Spence et al. (1979) and apparently consistent with Bem's original formulation, posits that M and F each contribute positively and uniquely to the prediction of self-concept. This concept implies that the main effects of both M and F will be statistically significant.

Masculinity Model. This model posits that self-concept, at least in modern Western societies, is primarily determined by M rather than F, and is sometimes called the "masculine supremacy effect" (Cook, 1985, p. 96). This model appears to be prompted primarily by empirical findings rather than by theory, though it may be consistent with feminist perspectives on the organization of society. Support for this model implies a main effect of M but no main effect of F, or perhaps a main effect of F in which F contributes negatively after the main effect of M is controlled for.

Interactive Androgyny Model. This model proposes that androgyny is more than the additive sum of M plus F. Thus, this model posits an M × F interaction.

Differentiated Additive Androgyny Model. This model is the first theoretical model of M/F and self-concept relations that incorporates the multidimensionality of self-concept. Like the additive model, this one posits that M and F each contribute positively to the prediction of self-concept. Its critical prediction, however, is that the relative contribution of M and F will differ systematically according to the specific facet of self-concept. As with the general construct validity approach applied in other SDQ research, it was posited that (a) M will be significantly more strongly related (for both males and females) to those facets of self-concept that are more logically related to M; and (b) F will be significantly more strongly related (for both males and females) to those facets of self-concept that are more logically related to F. In order to eliminate the possible circularity in this hypothesis (i.e., assuming facets that are more positively correlated to M or to F are more logically related to M or to F), facets most logically related to M are defined operationally to mean those facets of self-concept for which males have higher scores, and facets most logically related to F are defined operationally to mean those facets of self-concept for which females have higher scores.

Empirical Support for the Models

Empirical support (e.g., Antill & Cunningham, 1979, 1980), literature reviews, and meta-analyses (e.g., Cook, 1985; Spence, 1984; Taylor & Hall, 1982; Whitley, 1983) of the relationship between M/F and self-concept consistently provide the strongest support for the masculinity model. In perhaps the classic study, Antill and Cunningham (1979, 1980) related responses from five different M/F instruments and two different esteem measures and concluded, "In every case masculinity showed significant positive correlations with self-esteem in both sexes whereas the correlations with femininity were generally nil or negative" (1979, p. 783). In a reanalysis of this study, Marsh, Antill, and Cunningham (1987) demonstrated that the results provided clear support for the masculinity model, but these results provided no support for the androgyny, sex-typed, or interactive models. Meta-analyses by Taylor and Hall (1982) and by Whitley (1983) found that most of the variance in the relationship between M/F and self-concept could be explained by M whereas little contribution was made by the main effect of F or by M × F interactions, and that these findings were consistent across responses by males and by females. In her review Cook (1985) noted, "The best documented and robust association in the androgyny literature is that between masculinity and paper-and-pencil tests of self-esteem. Femininity is more weakly related, if at all" (p. 94). Hence, androgyny researchers have been unable to find much support for either the additive or the interactive models that are derived from and central to androgyny theory. This represents, perhaps, the most devastating appraisal of androgyny theory.

Most previous research, however, is seriously flawed (Marsh, 1987c). The most important limitation of this research is the exclusive use of undifferentiated, global self-concept measures. Thus, this previous research has been unable to test the differentiated additive androgyny model that is the focus of the present investigation.

Relation Between Masculine/Feminine Identity and SDQ-II

Using the SDQ-II, Marsh (1987c) investigated the relationship between masculine/feminine identity and differentiated aspects of self-concept. The purpose of the investigation was to examine the relations between 4 ASRS (positive characteristics M+ and F+, and negative characteristics M− and F−) scales and 11 facets of self-concept measured by the SDQ-II with respect to the theoretical models and issues discussed earlier. Marsh's major prediction was that the differentiated additive model would be supported in that (a) the relative contribution of M and F would vary according to the specific area of self-concept; (b) the contribution of M would be relatively larger for those facets for which males have higher self-concepts; and (c) the contribution of F would be relatively larger for those facets for which females have higher self-concepts.

The sample consisted of 962 (49% female, 51% male) high school students (grades 7–11) who attended one of two neighboring schools that served the same predominantly middle-class suburb of metropolitan Sydney, Australia. The two self-report instruments, the SDQ-II and the ASRS (Marsh, 1987c), were administered to students in each high school.

Analysis of Variance (ANOVA) Tests of the Models. Each of the five theoretical models makes different predictions that can be tested with an ANOVA. Most applications of the ANOVA approach have relied on a 2 × 2 design (Taylor & Hall, 1982; Whitley, 1983). The gross classification of M and F into dichotomies, however, may be dubious (Marsh, 1987c). The typical regression approach provides stronger tests of the theoretical models because M and F are not dichotomized. Spence (1984)

also recommended a 4 × 4 ANOVA design in which M and F are divided into quartiles. In the present investigation this 4 × 4 design is supplemented by regression analyses.

A general test of the differentiated additive androgyny model is based on a repeated-measures ANOVA. In addition to the effects of M, F, gender, and their interactions, each different facet of self-concept corresponds to one level of the repeated-measures variable that is called "scales" (self-concept) in the present investigation. For this analysis the model posits that the main effects of M and F will be statistically significant, that there will be a significant M-scale interaction (that the effect of M will vary with the facet of self-concept), that there will be a significant F-scale interaction (that the effect of F will vary with the facet of self-concept), and that no other interactions will be of practical significance. If these general predictions are supported, then further analyses are necessary to determine if the form of the M-scale and F-scale interactions are consistent with the model.

The sex-typed model posits that the effect of M and the effect of F (and perhaps their interaction) will vary according to gender. In particular, M should contribute more positively to self-concept for boys, and F should contribute more positively for girls. However, gender failed to interact significantly with M, with F, or with M × F. This lack of support for the sex-typed model was consistent across all 11 areas of self-concept. Given the statistical power of this test and the consistency of the findings across the 11 areas of self-concept, the lack of support for the sex-typed model was compelling.

The androgyny interactive models posit an M × F interaction, but the M × F interaction failed to reach statistical significance. This lack of M × F interaction was also consistent across all 11 areas of self-concept. Given the statistical power of this test and consistency of the findings across the 11 areas of self-concept, the lack of support for the interaction model was also compelling.

The additive androgyny model was supported for M/F groups based on positively valued items in that both M and F contributed substantially and the direction of the contribution was positive for all areas of self-concept. However, as predicted by the differentiated additive model, the substantial M-scale and F-scale interactions demonstrated that the size of the positive contributions of M and F varied significantly with the particular facet of self-concept. These interactions and their relation to the specific pattern of results posited in the differentiated additive model is the focus of the subsequent analyses.

Tests of the additive and differentiated additive models were more complicated for M/F groupings based on the negatively valued M/F items. However, the substantial M-scale and F-scale interactions in these analyses again demonstrated that the effects of both M and F varied significantly with the particular facet of self-concept, as predicted by the differentiated additive model.

The masculinity model posits that the main effect of M on self-concept is positive, whereas the effect of F is nonsignificant or negative. For the positively valued M/F items, there was no support for this model in that contributions of M and F were both positive. Support for this model was somewhat stronger for the negatively valued M/F items for which the contribution of M items was relatively more positive, or at least less negative, than the contribution of F. However, the relative contribution of M and of F varied substantially with the particular facet of self-concept, and this does not seem to be consistent with the masculinity model. Hence, support for the masculinity model is weak.

In summary, the ANOVA results provided no support for the androgyny interactive or sex-typed models. Support for the masculinity model was weak in that F,

particularly F+, contributed substantially to self-concept, and the relative contribution of M and of F varied with the area of self-concept. In contrast, support for the additive and the differentiated additive models was strong.

M/F Correlations with Multiple Dimensions of Self-Concept. Correlations and beta weights relating the 4 ASRS scores to the 11 SDQ-II scores are presented in Table 17. Consistent with the ANOVA results, these findings show that when averaged across all areas of self-concept (the mean of coefficients in Table 17), correlations are (a) modestly positive for the M+ and F+ positively valued scores, (b) smaller and negative for the F− score, and (c) close to zero for the M− score. Also consistent with the ANOVA results and with the differentiated additive androgyny model, the pattern of correlations between M/F and self-concept differs substantially depending on the area of self-concept.

The positively valued scores F+ and M+ both contribute positively to self-concept. In three areas—Honesty–Trustworthiness, Parent Relations, and Same-Sex Relations—the self-concept scores are more positively correlated to F+ than to M+, and two of these are areas of self-concept in which girls have substantially higher self-concepts than do boys (see Table 17). Boys have substantially higher self-concepts than do girls in Physical Appearance, Physical Abilities, and Math (Table 17), and these self-concept scores are more positively correlated with M+ than with F+.

Table 17. Correlations and Beta Weights for Four M/F Scores and SDQ-II Responses

	M/F Scores								
	M+		M−		F+		F−		
Self-Concepts	*r*	*beta*	*r*	*beta*	*r*	*beta*	*r*	*beta*	*Multiple R*
Physical Abilities	.33*	.38*	−.02	−.16*	.13*	−.01	−.18*	−.11*	.38*
Physical Appearance	.34*	.27*	.21*	.12*	.05	−.01	−.19*	−.17*	.38*
Opposite-Sex Relations	.34*	.23*	.12*	.08*	.17*	.14*	−.25*	−.26*	.42*
Same-Sex Relations	.21*	.05	.00	.03	.29*	.31*	−.14*	−.20*	.36*
Parent Relations	.09*	.13*	−.25*	−.28*	.22*	.16*	−.10*	−.08*	.35*
Honesty–Trustworthiness	.03	.09*	−.43*	−.43*	.36*	.29*	−.04	−.03	.55*
Emotional Stability	.27*	.25*	−.13*	−.16*	.09*	.07*	−.51*	−.47*	.58*
Math	.20*	.21*	.00	−.08*	.13*	.05	−.06	−.03	.22*
Verbal	.31*	.30*	−.04	−.14*	.25*	.15*	−.12*	−.10*	.39*
General School	.32*	.33*	−.02	−.14*	.21*	.09*	−.12*	−.08*	.37*
General Self	.46*	.42*	−.01	−.14*	.35*	.21*	−.19*	−.16*	.54*
Summary Statistics									
Mean Coefficient	.26	.24	−.09	−.12	.20	.13	−.17	−.15	.41
Mean of Squared Coefficients	.08	.07	.03	.04	.05	.03	.04	.04	.18

Note: N = 962 (471 females and 491 males). A series of 11 multiple regressions was conducted to relate the four M/F scores (M+, M−, F+, and F−) to the 11 self-concept scores. The simple correlations and standardized beta weights resulting from these analyses are presented in this table. Results for males and females were similar to each other and to those shown here.

*$p < .05$.

From "Masculinity, Femininity, and Androgyny: Their Relations to Multiple Dimensions of Self-Concept," by H. W. Marsh, 1987, *Multivariate Behavioral Research, 22*, p. 108. Copyright 1987 by Lawrence Erlbaum Associates, Inc. Adapted by permission of the publisher.

The correlations and beta weights in Table 17 also indicate that M⁻ and particularly F⁻ tend to contribute negatively to self-concept. However, M⁻ contributed positively to Opposite-Sex Relations, Physical Abilities, and Physical Appearance self-concepts favoring boys. F⁻ did not contribute positively to any areas of self-concept, but Honesty–Trustworthiness was the area in which the contribution of F⁻ was least negative (not significantly different from zero), the area of self-concept most favoring girls. The most negative contribution of F⁻ is for the Emotional Stability scale, but boys and girls did not differ significantly on this facet. An inspection of the content of the F⁻ items (e.g., anxious, nervous, worrying) suggests that this effect may be idiosyncratic to the ASRS. However, the F+ scale is also less positively correlated with the Emotional Stability scale than with other areas of self-concept, and Spence et al. (1979) reported F+ and F⁻ to be significantly related to neuroticism.

The correlations in Table 17 tend to support the differentiated additive model, but the large number of coefficients complicates the interpretations so that a more objective index is needed. The size and direction of the correlations between each self-concept and bipolar M/F scores based on positively valued items (M+ minus F+), on negatively valued items (M⁻ minus F⁻), and on their total (MTOT minus FTOT) provides such an index. The M scores are weighted +1, and the F scores are weighted ⁻1 in the computation of these bipolar scores. Thus, if the correlation between any bipolar M/F score and a self-concept score is positive, then M scores contribute more positively (or at least less negatively) than do F scores; if the correlation is negative, then the positive contribution of the F scores is larger; if the correlation does not differ significantly from zero, then the relative positive contribution of M and F scores is similar.

These three bipolar M/F scores were correlated with the self-concept scores (see Table 18). For each of the three bipolar M/F scores, 6 of 11 correlations were significantly positive, indicating a larger positive contribution for F than for M (2 did not differ significantly from zero). The consistency of this pattern of results for the positively and negatively valued M/F items offers support for the generality of the conclusions.

The differentiated additive model posits that F will contribute more positively to those areas of self-concept in which females have higher self-concepts and that M will contribute more positively in those areas in which males have higher self-concepts. The set of 11 correlations between the bipolar M/F scores and multiple self-concepts provides a quantitative index of the relative positive contribution of M and F to each area of self-concept. Similarly, correlations between gender (1 = male, 2 = female) and each area of self-concept provide a quantitative index of gender differences in each area of self-concept.

The relation between the two sets of correlations—correlations between the bipolar M/F scores and the self-concept scores and correlations between gender and the self-concept scores—provides a test of the differentiated additive model. This correlation is ⁻.81, ⁻.61, and ⁻.71 (df = 9, all p values $<.05$) for the positive, negative, and total M/F scores. Thus, the areas of self-concept most favoring girls (i.e., those in which correlations with gender are most positive) are the ones in which the positive contribution of F is larger than the contribution of M (i.e., correlations with bipolar M/F are most negative). These findings support the differentiated additive model: Areas in self-concept in which F contributed more positively than did M are the areas in which females have higher self-concepts than do males, and vice versa. The consistency of the results across both positively and negatively valued M/F items further supports the generality of the conclusions.

Table 18. Correlations Among Bipolar M/F Scores, Gender, and SDQ-II Responses

Variable	Bipolar M/F Scores For: Positive M/F Items	Negative M/F Items	Positive & Negative M/F Items	Gender[a]
SDQ-II Scales				
Physical Abilities	.16**	.10*	.14**	⁻.14**
Physical Appearance	.24**	.30**	.31**	⁻.34**
Opposite-Sex Relations	.14**	.27**	.24**	⁻.06
Same-Sex Relations	⁻.10**	.09**	.01	.29**
Parent Relations	⁻.12**	⁻.13**	⁻.15**	.00
Honesty–Trustworthiness	⁻.31**	⁻.32**	⁻.36**	.25**
Emotional Stability	.16**	.25**	.24**	⁻.05
Math	.05	.04	.05	⁻.12**
Verbal	.03	.05	.04	.07*
General School	.08**	.07*	.08**	.07*
General Self	.08*	.12**	.11**	⁻.05
Sex	⁻.37**	⁻.19**	⁻.30**	1.00

Note. N = 962 (471 females and 491 males).

[a]Gender is keyed so that 1 = male and 2 = female.

*$p < .05$; **$p < .01$.

Adapted from "Masculinity, Femininity, and Androgyny: Their Relation to Multiple Dimensions of Self-Concept," by H. W. Marsh, 1987, *Multivariate Behavioral Research, 22*, p. 110. Copyright 1987 by Lawrence Erlbaum Associates, Inc. Adapted by permission of the publisher.

Summary and Implications

Taken as a whole, the study by Marsh (1987c) provided no support for the interactive or sex-typed models, little support for the masculinity model, some support for the additive androgyny model, and very strong support for the differentiated additive androgyny model. The lack of support for the interactive and sex-typed models is particularly persuasive because of the large sample size and the use of the 4 × 4 design, because of the consistent lack of effect across the 11 facets of self-concept, and because the adolescent period is the period when the sex-typed effects are supposed to be strongest. In support of the additive model, both M and F, particularly M+ and F+ (positive scales of the ASRS), contributed positively to the prediction of self-concept. In support of the differentiated additive model, the relative contribution of M and F varied substantially and predictably with the specific area of self-concept. The contribution of F was more positive than that of M in the areas of self-concept where females had higher self-concepts, and the positive contribution of M was greater in those areas where males had higher self-concepts.

These findings make an important contribution to the study of the relationship between M/F and self-concept that is central to androgyny theory because (a) previous research has found little or no positive contribution of F after taking into account M, whereas F was shown to contribute substantially in the present study; (b) previous research has typically ignored the multidimensionality of self-concept, whereas the present investigation demonstrated that the relationship between M/F and self-concept varies substantially with the specific facet of self-concept; and (c)

the present investigation proposed and demonstrated strong empirical support for a new model of the M/F-versus-self-concept relationship, the differentiated additive androgyny model.

The results of this study also contribute further support to the construct validity of SDQ-II responses and demonstrate problems in relying on a single global indicator to infer self-concept. Here, as in many areas of research that incorporate measures of self-concept, seemingly confusing results based on global measures of self-concept become more easily understood when appropriately constructed multidimensional self-concept instruments, such as the SDQ, are used.

Chapter 10. Summary and Overview

Interest in self-concept stems from its recognition as a valued outcome in a wide spectrum of disciplines—from the assumption that the improvement of self-concept may facilitate improvements in other outcomes such as academic achievement, from interest in how self-concept is related to other variables, and from interest in particular measurement and methodological problems inherent in this area of research. The study of self-concept represents 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 on self-concept.)

There are also interesting peculiarities about research in this area. Unlike other areas of research, the study of self-concept is not aligned with any particular discipline. Also, although many thousands of self-concept studies have been conducted, only a few researchers have published a substantial number of studies or have continued their research over an extended period of time. In fact, most self-concept studies emphasize other theoretical constructs, and the interest in self-concept comes from its assumed relevance to these other constructs. Reviews of self-concept research (e.g., Burns, 1979; Shavelson et al., 1976; Wells & Marwell, 1976; Wylie, 1974, 1979) typically emphasize 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. To these can be added the apparently inappropriate overreliance on global indicators of self-concept to test the effects of interventions that are more logically related to specific components of self-concept. The disappointing lack of rigor can perhaps be explained by the lack of identification of self-concept research with any particular discipline.

In an attempt to clarify the status of the self-concept construct, Shavelson et al. (1976) reviewed existing theoretical and empirical research and developed their multifaceted, hierarchical model of self-concept. Although the facets proposed in the Shavelson et al. model, as well as their structure, were heuristic and plausible, they were not supported by empirical research. In particular, at the time of their review, Shavelson et al. were unable to identify any existing instruments that measured the facets posited in their model. Also, the purported multidimensionality of self-concept was not widely accepted by other researchers. Shavelson et al. speculated that the lack of empirical support for their model was due to the poor quality of measurement instruments employed in self-concept research. The development of a suitable instrument was clearly necessary for empirically testing Shavelson et al.'s model and for examining further theoretical issues.

This need provided the initial impetus for the development and refinement of the SDQ instruments, including SDQ-II. In adopting such an approach, atheoretical or purely empirical approaches to developing and refining the self-concept instrument were clearly 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 on which it was based. Implicit in this approach is the assumption 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 strong empirical and theoretical bases.

Because self-concept is a hypothetical construct, its usefulness must be demonstrated by investigations of its construct validity. These investigations can be classified as within-network or between-network studies. Within-network studies examine the dimensionality of self-concept. These studies typically employ factor analysis and multitrait–multimethod analysis to determine if self-concept is multidimensional and to identify the dimensions which characterize it. Between-network studies attempt to demonstrate a theoretically consistent, or at least logical, pattern of relationships between measures of self-concept and other constructs such as academic achievement. Research on the SDQ-II and its two companion instruments includes both within-network and between-network studies. Early research on these instruments, primarily within-network studies, focused on internal characteristics of self-concept, particularly facets of self-concept and the organization of these facets. More recent research, primarily between-network studies, examines the relationship between self-concept facets and a wide variety of other constructs including sex, age, academic performance, self-concepts inferred by significant others, family background characteristics, attributions for success and failure, and experimental interventions designed to enhance self-concept. Research described in this Manual provides strong support for the validity of interpretations based on responses to the SDQ-II and the Shavelson et al. model on which it is based. The research has also clarified many theoretical issues in self-concept research. In this sense this research represents an interplay between theory and empirical research and supports the construct validity not only of the SDQ-II but also of its two companion instruments—SDQ-I and SDQ-III.

Perhaps the strongest contribution of the accumulated research from SDQ-I, -II, and -III is the development of a carefully designed set of instruments, based on a strong empirical foundation and a good theoretical model, for the measurement of self-concept. Reviewers in this field typically argue that the most important weakness in self-concept research is the poor quality of measurement instruments. The development of the SDQ instruments, including SDQ-II, is intended to help remedy this situation.

Appendices

Chapter 7.
Self-Concept and Academic Achievement

Wylie (1979) noted that "many persons, especially educators, have unhesitatingly assumed that achievement and/or ability measures will be strongly related to self-conceptions of achievement and ability and to overall self-regard as well" (p. 355). Not surprisingly, particularly for studies of school-aged children, some measure of academic achievement is one of the criteria most frequently used to validate self-concept interpretations. Academic achievement has also been the focus of much SDQ research. This research is reviewed more extensively in the SDQ-I and SDQ-III manuals (also see Marsh, 1986d) but is summarized briefly in this chapter.

The SDQ and the Shavelson et al. model upon which it is based hold self-concept to be a multifaceted, hierarchically ordered construct. Academic self-concept is one component of overall self-concept that is separated into particular content areas such as reading and math. Support for the construct validity of SDQ interpretations and the Shavelson et al. model requires (a) academic achievement to be more positively correlated with academic self-concept than with indices of nonacademic or general self-esteem, and (b) verbal and math achievement to be more highly correlated with self-concepts in matching content areas than with other facets of self-concept. In a meta-analysis of the achievement–self-concept relationship, Hansford and Hattie (1982) found that measures of ability–performance correlated about .20 with measures of general self-concept but also correlated about .40 with measures of academic self-concept. This and other research reviewed in the SDQ-I Manual underscore the importance of distinguishing among academic, nonacademic, and general self-concepts.

Research based on all three SDQ instruments has generally supported this relationship between specific facets of academic self-concept and corresponding measures of academic achievement, such as test scores or grades. For example, math achievement is most highly correlated with Math self-concept, English achievement is most highly correlated with Reading/Verbal self-concept, and so forth. Support for the generality of this pattern of results comes from several studies (Marsh, Parker, & Barnes, 1985; Marsh & Peart, 1988; Thomas, 1984; also see Marsh, 1986c).

Relationship of Self-Concept to Academic and Other Constructs

Educational Coping Skills

Thomas (1984), as part of a thesis on educational coping skills, administered the SDQ-II in a battery of pretests. Subjects were ninth grade students attending a boys' (N = 79) or girls' (N = 51) high school in metropolitan Sydney, Australia. In addition to the SDQ-II, the study also included several academically oriented variables.

Mathematics Ability Grouping. School administrators classified students into one of four *streams* (i.e., ability groupings) based on their mathematics achievement and performance in previous mathematics classes. For purposes of this study, these mathematics ability groupings were taken as measures of academic performance. Students assigned to the lowest ability level were given a score of 1, students in the next lowest ability level were given a 2, and so forth.

Study Skills Inventory. Responses to 175 items were used to assess study skills in seven areas (place of study, study times, organization, textbook reading skills, note taking, studying for exams, and exam technique). Only the total score was used in the present analysis.

Anxiety Measures. Students also responded to two anxiety measures. The *Test Anxiety Inventory* (Spielberger, Gonzalez, Taylor, Anton, Algaze, Ross, & Westberry, 1980) is designed to measure individual differences in test anxiety as a situation-specific trait. The measure contains separate scales for worry and emotionality. The second measure, the *State–Trait Anxiety Inventory*, is a general measure of anxiety proneness or trait anxiety (Spielberger, Gorsuch, & Luchene, 1970).

Survey of Study Processes. Biggs (1978, 1984) devised this survey to measure three levels of academic motivation called *surface, deep,* and *achieving*. In the *surface* dimension the motive for study is a means to an external need (e.g., a job). The strategies are to avoid failure, to focus on minimal requirements, and to learn by rote. In the *deep* dimension the motives are to develop specific interests and to maximize intrinsic motivation. The strategy is to read widely for maximum understanding, often independently of specific course requirements. In the *achieving* dimension the motives are to excel in a competitive environment, to achieve high status, and to maximize the grades awarded. The strategies are to follow carefully the course outline, to complete assignments on time, and so forth.

Correlations between the SDQ-II responses and the set of nine criterion variables are presented in Table 10. All but one of the criterion variables, General Anxiety, are academic constructs and thus should be more highly correlated with the academic SDQ-II scales than with the nonacademic SDQ-II scales. Generally, support for this predicted pattern of relations is good which, in turn, supports the construct validity of the SDQ-II responses. Even though the Test Anxiety and General Anxiety measures are substantially correlated (.62), Test Anxiety is more highly correlated with the academic self-concept scales, whereas General Anxiety is more highly correlated with General self-concept (in particular, Emotional Stability) than with any of the academic self-concepts. Significant correlations with study process variables and Same-Sex Relations, although generally lower than those for the academic self-concept scales, suggest that peer interactions may be relevant in study processes and ability groupings. Overall the results suggest that students with high academic self-concepts are more likely to be in advanced ability groupings, to have good study skills, to be less anxious in testing situations, and to be more highly motivated to study for deep understanding and achievement than to study for superficial understanding that meets minimal requirements.

Table 10. Correlations Between the SDQ-II Scales and Selected Academically Related Variables and Anxiety

SDQ–II Scales	Math Ability Grouping	Study Skills	Study Processes			Test Anxiety			General Anxiety
			Surface	Deep	Achieving	Total	Worry	Emotionality	
Physical Abilities	.12	.29**	.01	.24**	.21*	.03	-.01	.06	-.24**
Physical Appearance	.05	.05	.10	.09	.05	.00	.05	-.01	-.16*
Opposite-Sex Relations	.07	.12	.05	.09	.09	.11	.12	.12	-.13
Same-Sex Relations	.22*	.52**	.06	.15*	.31**	-.07	-.15	.01	-.36**
Parent Relations	.11	.29**	.06	.21*	.20*	-.14	-.14	-.08	-.34**
Honesty–Trustworthiness	.22*	.50**	-.09	.15*	.24**	-.13	-.27**	-.02	-.35**
Emotional Stability	.13	.30**	.06	.24**	.15*	-.20*	-.20*	-.14	-.55**
Math	.48**	.42**	-.25**	.20**	.39**	-.34**	-.39**	-.27**	-.33**
Verbal	.39**	.62**	-.12	.37**	.45**	-.11	-.25**	.00	-.29**
General School	.55**	.69**	-.10	.35**	.52**	-.22**	-.37**	-.07	-.40**
General Self	.31**	.46**	-.05	.17*	.17*	-.17*	-.24**	-.07	-.51**

Note. $N = 130$, 79 males and 51 females. The criterion measures are scored such that positive correlations mean that high self-concepts are associated with more advanced ability grouping (math stream), more effective study skills, stronger orientations toward each study process orientation, and lower levels of anxiety.
$^{*}p < .05$; $^{**}p < .01$.

From "An Investigation into the Influence of Anxiety and the Effectiveness of Treatments Involving Positive Coping Skills to Alleviate Anxiety's Detrimental Effects," by T. A. Thomas, 1984, unpublished bachelor's honors thesis, University of Sydney, Australia.

Math Achievement and Physical Fitness

In another study, Marsh and Peart (1988; also see Chapter 8) tested the effects of physical fitness training programs on measures of academic and nonacademic self-concept. The sample included 137 eighth grade girls who attended a Catholic girls' school in metropolitan Sydney, Australia. Pretest measures in this study included responses to the SDQ-II and performance on a comprehensive set of seven physical exercises (running speed, number of push-ups, etc.) that were combined to form a total physical fitness measure. Though not a focus of the study, the math ability grouping for each girl was also recorded. This ability grouping was one of five levels to which the student was assigned by the school and was based on objective test scores and previous school performance.

Correlations between the SDQ-II scales and the two criterion measures are presented in Table 11. The mathematics ability grouping is substantially correlated with Math (.48) and General School (.50) self-concepts, less correlated with Verbal self-concept (.27), and even less correlated with the remaining nonacademic dimensions of self-concept (⁻.16 to .23). In addition, Math self-concept is substantially correlated with math achievement (.48) and not significantly related to physical fitness (⁻.09). Physical fitness, on the other hand, is substantially correlated with Physical Abilities self-concept (.45) but not with any of the remaining SDQ-II scales (⁻.09 to .11). Furthermore, Physical Abilities self-concept is substantially correlated with physical fitness (.45) but is not significantly correlated with mathematics achievement (⁻.08). Math achievement and the total physical fitness index were uncorrelated (.01). The juxtaposition of these two sets of correlations provides strong support for the convergent and divergent validity of the SDQ-II responses, as they relate to these two specific facets of self-concept.

Table 11. Correlations of SDQ-II Scales with Measures of Math Achievement and Physical Fitness

Scale	Coefficient Alpha	Math Achievement	Physical Fitness
Physical Abilities	.89	⁻.08	.45*
Physical Appearance	.91	.13	.07
Opposite-Sex Relations	.93	⁻.16	.02
Same-Sex Relations	.90	.09	.00
Parent Relations	.90	.22*	⁻.07
Honesty–Trustworthiness	.77	.23*	⁻.09
Emotional Stability	.87	.12	.11
Math	.92	.48*	⁻.09
Verbal	.88	.27*	.00
General School	.89	.50*	⁻.08
General Self	.91	.23*	.07

Note: $N = 137$.

*$p < .01$.

Adapted from "Competitive and Cooperative Physical Fitness Training Programs for Girls: Effects on Physical Fitness and on Multidimensional Self-Concepts," by H. W. Marsh, & N. D. Peart, 1988, *Journal of Sport Psychology, 10*, p. 399. Copyright 1988 by Human Kinetics Publishers. Adapted by permission of the publisher.

Measures of Academic Achievement

In another study, Marsh, Parker, and Barnes (1985) investigated the relationship between self-reported self-concept, as assessed by the SDQ-II, and a variety of academic achievement measures. The sample consisted of 901 students (47% females, 53% males) in grades 7 through 12 who attended the same high school in metropolitan Sydney, Australia. The students as a group had an average IQ range (M = 101.6, SD = 13.4), although students in grades 11 and 12 showed somewhat above-average intellectual ability. A variety of academic criterion measures was obtained for each student from the school records. A brief summary of each of these follows.

General Achievement. All students in the school district were administered a standardized achievement test during the sixth grade.

Reading Achievement. Students, upon entering the high school, were given a standardized reading test. Because this procedure had been only recently instituted, these scores were available only for the seventh and eighth grade students.

Mathematics and English Ability Groupings. In the seventh grade, the first year of high school in Australia, students were classified into ability groupings based on their general achievement scores. For this first year only, students were assigned to the same level in all subjects. In subsequent years, students were assigned to separate ability groupings in mathematics and English depending on their school performance in each subject area. The number of ability groupings varied (10 each in grades 7, 8, and 9; 7 in grade 10; and 3 in grades 11 and 12) as a function of the number of students and staffing availability. For purposes of this study, the mathematics and English groupings (and reading achievement test scores in grades 7 and 8) were taken as measures of academic performance. As in the study by Thomas (1984) cited previously, students assigned to the lowest ability level were given a score of 1, students in the next lowest ability level were given a score of 2, and so forth.

The general achievement scores were the only academic achievement measures on which students from different grade levels could be compared. Preliminary analyses showed that students in grades 11 and 12 had significantly higher scores (M = 110.6) than did other students, but that differences between other years (means ranged from 99.6 in grade 7 to 102.2 in grade 10) and between 11th and 12th graders were not statistically significant. This pattern is consistent with the self-selection that takes place after year 10 in Australian public schools and supports the decision to combine students from years 11 and 12 into a single group. It also indicates the need for caution in interpreting age effects beyond year 10 because the students in grades 11 and 12 are not comparable with the rest of the sample.

The relationship between the academic criteria and self-concept scales was examined separately for each grade level (see Table 12) because the criteria and their interpretation varied for the different grade levels. Nevertheless, some generalizations can be made across the grade levels. At all five grade levels, each of the academic self-concepts and their total are significantly correlated with the set of academic criterion measures. For each grade level, multiple correlation was used to predict each self-concept score from all available indicators of academic achievement. The 20 multiple correlations (multiple R) relating achievement indicators to academic self-concepts (Math, Verbal, General School, and Total Academic) vary from .20 to .55 (median multiple R = .38). Of the 40 multiple Rs relating the academic criterion to nonacademic scales, only 8 reach statistical significance, and none is larger than .30 (median = .18). This pattern shows the specificity of the

Table 12. Correlations Among Self-Concept Scores, Academic Achievement, and IQ

Variable	Physical Abilities	Physical Appearance	Opposite-Sex Relations	Same-Sex Relations	Parent Relations	Self-Concept Scales: Honesty–Trustworthiness	Emotional Stability	Math	Verbal	General School	General Self	Total Academic	Total Self-Concept
Grade 7													
IQ	.03	.03	−.16	.09	.08	.17	.22	.05	.28	.18	.05	.24	.19
English	.10	.05	−.14	.11	.11	.20	.23	.13	.25	.21	.11	.28	.26
Reading	−.04	−.01	−.20	.06	.10	.17	.14	.12	.34	.19	.09	.30	.17
Multiple R	.22*	.11	.21*	.12	.11	.20*	.23**	.20*	.35**	.21*	.17	.31**	.27**
Grade 8													
IQ	−.06	.11	−.02	.23	−.06	.06	.23	.07	.14	.10	−.04	.14	.14
Math	.02	.11	−.07	.24	−.01	.14	.18	.18	.13	.16	−.01	.21	.19
English	.01	.05	−.09	.23	−.01	.14	.19	.10	.20	.13	−.03	.19	.16
Reading	−.13	.06	−.04	.16	−.04	.12	.18	.06	.21	.05	−.04	.14	.11
Multiple R	.24*	.25**	.16	.24**	.12	.20	.23*	.30**	.31**	.21*	.08	.23*	.20
Grade 9													
IQ	−.09	−.05	−.11	.08	−.17	−.08	.15	.07	.30	.25	.02	.29	.07
Math	−.08	−.02	−.13	.03	−.07	.08	.05	.27	.15	.35	.06	.36	.13
English	−.07	−.05	−.12	.12	−.09	.03	−.03	.08	.25	.28	.05	.29	.08
Multiple R	.09	.09	.13	.18	.18	.20	.30**	.41**	.34**	.36**	.14	.38**	.14
Grade 10													
IQ	−.06	.06	−.04	−.02	.06	.07	.14	.18	.27	.21	−.01	.31	.16
Math	.03	.03	−.08	.05	.03	.10	.08	.45	.10	.32	−.03	.41	.21
English	.01	.04	−.08	.13	.12	.15	.05	.26	.26	.31	.02	.38	.24
Multiple R	.13	.07	.10	.24*	.20	.17	.18	.55**	.39**	.33**	.10	.41**	.24*
Grades 11/12													
IQ	.11	.07	−.02	.17	−.09	−.02	.24	.14	.35	.27	.00	.35	.22
Math	.00	.10	−.07	.21	.05	−.11	.19	.34	.14	.32	.17	.38	.25
English	.14	.24	.09	.26	.03	−.07	.11	.20	.35	.22	.15	.17	.23
Multiple R	.18	.25	.15	.28	.17	.13	.25	.53**	.45**	.34*	.29	.41**	.29

Note: $N = 901$. The Total Academic Scale is the sum of the means of the Math, Verbal, and General School scales.

$^*p < .05$; $^{**}p < .01$.

From "Multidimensional Adolescent Self Concepts: Their Relationship to Age, Sex and Academic Measures," by H. W. Marsh, J. Parker, & J. Barnes, 1985, *American Educational Research Journal, 22*, p. 437. Copyright 1985 by the American Educational Research Association. Adapted by permission of the publisher.

self-concept ratings in that academic criteria are related primarily to academic self-concepts and not to the nonacademic areas of self-concept.

Evidence is also strong for the specificity of self-concept within the different academic areas. Although General School self-concept is moderately correlated with both the Math and the Verbal self-concept scales, the Math and Verbal scales are not significantly correlated with each other in the total sample or at any grade level. In grades 7 and 8, the only two grades where a standardized reading achievement test score was available, reading achievement is significantly correlated with Verbal self-concept ($r = .34$ and .21, respectively, $ps < .01$) but not with Math self-concept ($r = .12$ and .06). Similarly, for grades where mathematics and English classes were streamed separately, the mathematics ability grouping is most highly correlated with Math self-concept, and the English ability grouping is most correlated with Verbal self-concept. Similarly, mathematics achievement scores are more highly correlated with Math self-concepts than with Verbal self-concepts, whereas verbal achievement scores are more highly correlated with Verbal self-concepts. However, on a technical note, this degree of specificity is severely limited by the extent of co-linearity between the verbal and math criterion scores. The correlation between the math and English ability grouping scores is 1.0 in grade 7 (where subjects are not streamed separately), .95, .89, and .89 in Grades 8, 9, and 10, and .47 in the combined grades 11 and 12 group.

To examine further this specificity, Marsh, Parker, and Barnes performed a new set of multiple regressions was performed that included only the math and English ability groupings as predictors of each of the academic self-concept scores. As before, analyses were computed separately for each year group. Every multiple R (see Table 13) is statistically significant and nearly as large as in Table 12, where multiple Rs were based on IQ scores as well as on the math and English ability grouping scores considered here. In Table 13 the standardized beta weights for each of the ability grouping scores are shown instead of the first-order correlations shown in Table 12.

Table 13. Standardized Beta Weights from the Multiple Regressions Relating Academic Self-Concepts to Academic Achievement

Variable	Academic Self-Concept Scores: Math	Verbal	General School	Total Academic
Grade 8				
Math	.86**	−.58**	.43*	.34
English	−.71**	.75**	−.28	−.13
Multiple R	.29**	.27**	.19*	.22**
Grade 9				
Math	.81**	−.27*	.44**	.46**
English	−.62**	.49**	−.11	−.11
Multiple R	.41**	.34**	.37**	.38**
Grade 10				
Math	1.03**	−.54**	.13	.34*
English	−.66**	.73**	.21	.08
Multiple R	.54**	.35**	.32**	.41**
Grades 11/12				
Math	.55**	.03	.28*	.38**
English	−.45**	.37**	.09	.00
Multiple R	.53**	.36**	.33*	.38**

$^*p < .05$; $^{**}p < .01$.

From "Multidimensional Adolescent Self-Concepts: Their Relationship to Age, Sex, and Academic Achievement," by H. W. Marsh, J. Parker, & J. Barnes, 1985, *American Educational Research Journal, 22*, p. 439. Copyright 1985 by American Educational Research Association. Adapted by permission of the publisher.

For Math self-concepts at every grade level, the direction of the math ability grouping score is significant and positive, whereas the direction of the English ability grouping is significant and *negative*. Once the effect of math achievement is controlled, a higher English achievement is associated with a lower Math self-concept. Conversely, for Verbal self-concepts, the direction of the English ability grouping score is positive, whereas the direction of the mathematics ability grouping score is negative. For the General School and Total Academic scores, there is no instance when both the math and English ability grouping scores add significantly to the multiple *R*, let alone have beta weights that are both significant and in the opposite direction. The extent of the specificity of this relationship between academic self-concept and academic ability is quite remarkable, given the large correlations between the English and math ability levels. These findings also demonstrate the clear separation between Verbal and Math self-concepts, even when the abilities in these two areas are substantially correlated.

Internal–External (I–E) Frame of Reference Model

Relationship of Verbal and Math Self-Concepts

Two sets of findings observed with the SDQ-II responses and with other SDQ research led to the development of the internal–external (I–E) frame of reference model. First, contrary to expectations based on the original Shavelson et al. model, Verbal and Math self-concepts were nearly uncorrelated (see Marsh & Shavelson, 1985, for details). Second, math achievement was negatively correlated with Verbal self-concept (after verbal achievement was controlled for), and verbal achievement was negatively correlated with Math self-concept (after math achievement was controlled for).

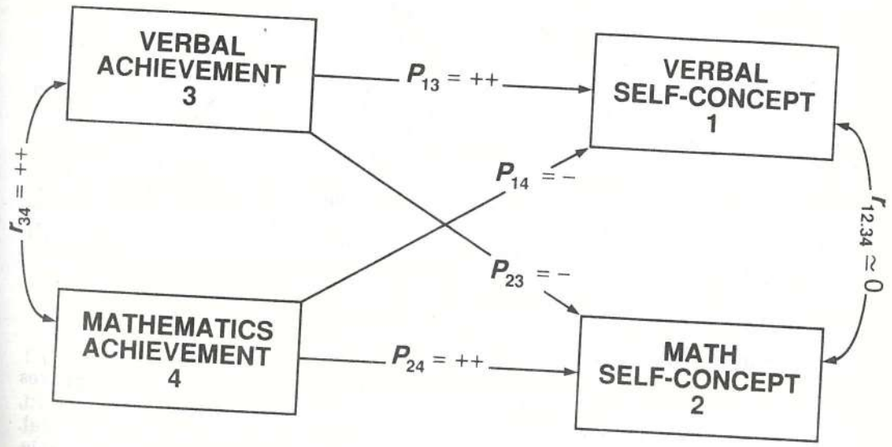
The I–E ("frame of reference") model was tested by several studies that examined the relationships between Verbal and Math self-concepts, and between these academic self-concepts and verbal and math achievement. According to the I–E model, Math and Verbal self-concepts are formed in relation to both external and internal comparisons, or frames of reference. In the external process, students compare their self-perceptions of their own math and verbal achievements with the perceived abilities of other students in their frame of reference and use this external, relativistic impression as one basis of their academic self-concepts. The model also posits an internal process by which students compare their perceived math achievement with their perceived verbal achievement and use this internal comparison as a second basis for forming their academic self-concept in each of these areas.

The external process has been well documented in self-concept research (e.g., Marsh & Parker, 1984) and more generally as a social comparison process (Suls & Miller, 1977). Because verbal and math achievements are substantially correlated, this external comparison process should lead to a positive correlation between Verbal and Math self-concept scores as originally anticipated in the Shavelson et al. model.

The internal comparison process, though more unusual in other theoretical accounts, is like the compensatory model described by Byrne (1984). Because math achievement and verbal achievement are compared with each other, and it is the *difference* between the two that contributes to a high self-concept in one area or the other, the internal process should lead to a negative correlation between Verbal and Math self-concept scores. The joint operation of both processes, depending on the

relative strength of each, may help to explain the near-zero correlations that have been observed in empirical research. The I–E model also predicts a *negative* direct effect of math achievement on Verbal self-concept, and of verbal achievement on Math self-concept. For example, a high Math self-concept will be more likely when math achievement is good (the external comparison) *and* when math achievement is better than verbal achievement (the internal comparison). Thus, once math skills are controlled, it is the *difference* between math achievement and verbal achievement that is predictive of Math self-concept; high verbal achievement, according to this formulation, will actually detract from a high Math self-concept. Thus, according to the I–E model, having good mathematics skills detracts from Verbal self-concept, and having good verbal skills detracts from Math self-concept.

Figure 15 summarizes the predicted relations among verbal achievement, math achievement, Verbal self-concept, and Math self-concept. In this model, academic achievements are hypothesized to be one causal determinant of academic self-concepts. However, this hypothesis does not argue against a more dynamic model where subsequent levels of academic achievement and self-concept are each determined by prior levels of achievement and self-concept. According to the path model, verbal and math achievements are substantially correlated with each other, whereas Verbal and Math self-concepts are nearly uncorrelated. Verbal achievement has a strong, positive, direct effect on Verbal self-concept but a small, negative, direct effect on Math self-concept. Similarly, math achievement has a strong positive effect on Math self-concept but a weaker, negative effect on Verbal self-concept (see Marsh, 1986d, for further discussion).



Note. Coefficients indicated to be "++", "−", and "0" are predicted to be high positive, low negative, and approximately zero, respectively.

From Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal, 23*, p. 134. Copyright 1986 by the American Educational Research Association. Reprinted by permission of the publisher.

Figure 15. Path Model of Relationships Among Achievements and Reading and Math Self-Concepts: Internal–External Frame of Reference Model

Empirical Support for the I–E Model

First, partial support for the I–E model examined by Marsh (1986d) was the near-zero correlation between Math and Verbal self-concepts. Correlations between Math and Verbal self-concepts were reported, based on responses to the SDQ-I, -II, and -III from 11 different studies which spanned ages from preadolescence to young adulthood. Except for the youngest children, those in second and third grades, the correlations were consistently close to zero. One of these studies was the SDQ-II study by Marsh, Parker, and Barnes (1985). Correlations between Math and Verbal self-concepts were not significantly different from zero in any of the five high school years considered (⁻.17 to .08). Similarly, correlations between Math and Verbal self-concepts were nonsignificant for the combined total sample (.00), and for the samples of boys (.07) and girls (⁻.02), respectively. These findings are also consistent with the factor analytic results based on the entire normative data sample of SDQ-II responses described in Chapter 3.

Second, the I–E model is further supported by tests of the predictions from the model illustrated in Figure 15 with data from six of the studies that also had math and verbal achievement scores. These results demonstrate that the direct effects of math achievement on Verbal self-concept and of verbal achievement on Math self-concept are consistently negative.

For example, consider the study by Marsh, Parker, and Barnes (1985) summarized previously in Table 13 (p. 65). As predicted by the I–E model, the direct effect of math achievement on Math self-concept is substantial and positive (βs = .55 to 1.03), whereas the direct effect of English achievement is smaller and negative (βs = ⁻.45 to ⁻.71). For Verbal self-concept, the direct effect of English achievement is large and substantial (βs = .37 to .75), whereas the direct effect of math achievement is smaller and negative for three of the four comparisons (βs = ⁻.27 to ⁻.58). In one of the four comparisons (grade 11/12), the contribution of math achievement to Verbal self-concept was not statistically significant, but this could be attributed to the fact that the students in this sample were highly self-selected.

In summary, the results of the Marsh, Parker, and Barnes (1985) study provide clear support for the I–E model. This support is consistent with the results of other research used to test the I–E model with responses to the SDQ-I and SDQ-III instruments.

Generality of the I–E Model

Marsh (1986d) noted that support for the I–E model was based entirely on responses by Australian students to one of the SDQ instruments. Following this observation, Marsh, Byrne, and Shavelson (1988) tested the generalizability of the I–E model. Nearly 1,000 Canadian students, 11th and 12th graders from two high schools, responded to the Verbal, Math, General School, and General Self scales from three different self-concept instruments. Although the results of this study were based in part on SDQ-III responses instead of on SDQ-II responses, the results are summarized here because they are based on responses by high school students and because they provide such an important test of the generality of results found for the SDQ-II.

Verbal, Math, and General School self-concepts were each measured with their respective scales from the SDQ-III, the *Self-Concept of Ability Scale* (SCA; Brookover, 1962; Shavelson & Bolus, 1982), and the *Affective Perception Inventory* (API; Soares & Soares, 1979). General esteem was measured by the SDQ-III General Self scale and by comparable scales on the API and Rosenberg's (1965) *Self-Esteem Scale* (SES). School grades for mathematics, English, and all school subjects were used as

indicators of math achievement, verbal achievement, and school achievement, respectively. In the initial analyses the I–E model was tested with responses from each of the self-concept instruments separately and with combined self-concept scores derived from all the instruments. These analyses tested the generality of the I–E model to non-Australian respondents and to different self-concept instruments. The I–E model posits that the correlation between Verbal self-concept and Math self-concept will be small and substantially smaller than the correlation between verbal achievement and math achievement.

Table 14 presents the correlation matrix among the 12 self-concept scores (Verbal, Math, General School, and General Self for the three instruments) plus combined scores based on all three instruments. Correlations with achievement (Math, Verbal, and General) and sex are also provided. The correlations between Verbal self-concept and Math self-concept varied between ⁻.05 and .08 for the different instruments and was zero for the combined scores. Despite this lack of correlation between Verbal self-concept and Math self-concept, verbal achievement and math achievement were substantially correlated (.51). Hence, the replicability and generality of this first prediction were supported.

The I–E model further predicts that verbal achievement will have a substantial positive effect on Verbal self-concept but a smaller negative effect on Math self-concept, whereas math achievement will have a substantial positive effect on Math self-concept but a smaller negative effect on Verbal self-concept. Support for this predicted pattern of results was also demonstrated by path analyses of scores based on each instrument separately and for the combined scores based on all three self-concept instruments (see Figure 16A). Hence, the replicability and generality of this second prediction were also supported.

Because of the substantial correlations among the achievement measures, the content specificity of achievement/self-concept relations was not so evident in the zero-order correlations. However, the results of the path analyses used to test the I–E model were ideally suited for making such comparisons; each path coefficient relating any one achievement score to any one of the self-concept scores represents the effect of that achievement score after the effect of the other achievement scores was partialled out. Math self-concept was substantially and positively related with math achievement, modestly and negatively related with verbal achievement, and nearly unrelated with school achievement. Verbal self-concept was substantially and positively related with verbal achievement, modestly and negatively related with math achievement, and relatively uncorrelated with school achievement. Academic self-concept was substantially correlated with school achievement and nearly uncorrelated with either verbal or math achievement. This pattern was consistent for self-concept scores from each instrument considered separately and for the combined scores.

In order to further test the generality of predictions based on the I–E model, Marsh, Byrne, and Shavelson (1988) added other variables to the analyses: General Self-Concept measures based on the three different instruments (Figures 16B, C, & D); General School self-concepts based on the three different instruments (Figures 16A–D); and a total academic achievement measure based on grades in all courses (Figures 16C & D). General self-concept, for each of the three instruments considered separately and for the combined measure, was not significantly correlated with academic achievement in English, mathematics, or all school subjects. The consistency of this finding across the three instruments and the large sample size demonstrate dramatically that general self-concept or self-esteem is relatively unaffected by academic achievement and that achievement is relatively unaffected by general self-concept. From this perspective, general self-concept may not be a particularly useful construct in educational research.

Table 14. Correlations Among Variables Considered in Tests of the I–E Model

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Self-Description Questionnaire-III (SDQ-III)																				
1 General Self	(.94)																			
2 General School	.19	(.90)																		
3 Verbal	.22	.37	(.80)																	
4 Math	.12	.36	−.05	(.93)																
Affective Perceptions Inventory (API)																				
5 General Self	.62	.20	.18	.16	(.86)															
6 General School	.55	.59	.37	.31	.62	(.85)														
7 Verbal	.06	.41	.68	−.05	.09	.42	(.89)													
8 Math	.10	.41	.02	.86	.16	.34	.08	(.95)												
Self-Esteem Scale (SES) and Self-Concept of Ability Scale (SCA)																				
9 General Self	.79	.24	.26	.19	.62	.59	.12	.19	(.87)											
10 General School	.17	.65	.32	.44	.17	.54	.33	.47	.23	(.87)										
11 Verbal	.04	.43	.54	−.01	.06	.33	.65	.03	.11	.49	(.90)									
12 Math	.07	.34	−.04	.82	.13	.31	−.04	.80	.12	.50	.03	(.94)								
Combined																				
13 General Self	.91	.24	.25	.18	.84	.66	.10	.17	.91	.21	.08	.12	(.94)							
14 General School	.35	.87	.41	.43	.39	.83	.46	.48	.41	.86	.49	.45	.43	(.93)						
15 Verbal	.13	.46	.85	−.04	.13	.44	.90	.05	.19	.44	.84	−.02	.17	.52	(.94)					
16 Math	.10	.39	−.02	.95	.16	.34	−.01	.94	.18	.50	.02	.93	.17	.48	.00	(.97)				
Achievement																				
17 School	−.02	.53	.16	.38	−.01	.36	.15	.32	.05	.69	.33	.41	.07	.62	.25	.40	(—)			
18 Verbal	−.04	.47	.24	.20	−.04	.32	.29	.16	.03	.54	.50	.18	−.02	.52	.40	.19	.80	(—)		
19 Math	−.04	.34	.02	.55	−.02	.26	.02	.49	.01	.52	.11	.62	−.02	.44	.06	.59	.74	.51	(—)	
20 Sex	−.04	.20	.12	−.17	−.17	.15	.25	−.17	−.11	.09	.19	−.16	−.12	.17	.22	−.18	.19	.26	.10	(—)

Note: Correlations greater than .07 are statistically significant ($p < .05$). Reliability estimates, coefficient alphas based on standardized responses to individual items, appear in parentheses. Because the achievement scores and sex were each measured by a single variable, coefficient alphas could not be computed (1 = male, 2 = female).

From "A Multifaceted Academic Self-Concept: Its Hierarchical Structure and Its Relation to Academic Achievement," by H. W. Marsh, B. M. Byrne, & R. Shavelson, 1988, *Journal of Educational Psychology, 80*, p. 372. Copyright 1988 by the American Psychological Association. Adapted by permission of the publisher.

The inclusion of gender (Figures 16B & 16D) is important for several reasons. First, the support for the I–E model may, in some way, be an artifact of gender-related differences in Math and Verbal self-concepts. The results in Figures 16 A–D, however, show that controlling for gender (by making it the first variable in path model) has little effect on the results. Support for the I–E model is still strong even after gender is controlled.

Second, the effect of gender on different areas of self-concept is substantively interesting (see discussion in Chapter 6), and the results presented here offer an important perspective. Girls have higher levels of verbal achievement and Verbal self-concept than do boys. Controlling gender differences in Verbal self-concept for gender differences in verbal achievement, however, does not eliminate the gender differences in Verbal self-concept. Thus, the advantage that girls have over boys in Verbal self-concept is more than can be explained in terms of differences in verbal achievement levels.

A different pattern of results exists for gender differences and Math self-concepts. Whereas boys have higher Math self-concepts than do girls, girls have higher levels of math achievement than do boys. Thus, the advantage that boys have over girls in Math self-concept cannot be explained at all in terms of math achievement (school grades) in this study. In fact, controlling for mathematics achievement actually increases—rather than decreases—sex differences in Math self-concept.

In both instances—higher Verbal self-concepts for girls and higher Math self-concepts for boys—gender differences in these academic areas of self-concept are consistent with sex stereotypes (see Marsh, 1989b; Marsh, Byrne, & Shavelson, 1988, for additional discussion). Furthermore, this pattern of results was consistent across the three different self-concept instruments.

In summary, Marsh, Byrne, and Shavelson's (1988) findings summarized here provide strong support for the multidimensionality of self-concept and the content specificity of general self-esteem, Verbal self-concept, Math self-concept, and academic self-concept. In particular, the path analytic results suggest that (a) general self-esteem is unaffected by verbal achievement, math achievement, or school achievement, (b) only verbal achievement has a positive influence on Verbal self-concept, (c) only math achievement has a positive influence on Math self-concept, and (d) only school achievement has a positive influence on General School self-concept. Other researchers have argued for the content specificity of different facets of academic self-concept (e.g., Marsh & Shavelson, 1985). However, no other research provides such strong support for the specificity of the different academic facets or the generality of this specificity across different self-concept instruments.

Studies of the I–E model also have practical implications for educators at all levels. An important dilemma faced by teachers is how to give positive feedback and praise that is realistic and honest and that will be accepted by academically poor students. If teachers are able to infer more accurately the academic self-concepts of their students and better understand how they are formed, then their ability to provide positive reinforcement to students of all ability levels will be enhanced. Even though teachers are able to infer student self-concepts in academic areas with at least modest accuracy, there appear to be several biases in their inferences. It is unjustified to assume that an academically weak student will necessarily have poor academic self-concepts in all settings and in all subject areas. First, students in settings where other students are also academically weak will have higher academic self-concepts than they would in settings where other students are academically average or above average. Previous SDQ research (Marsh, 1986d) suggests that teachers emphasize absolute measures of academic achievement when inferring the academic self-concepts of their students and largely ignore the particular setting that establishes the frame of reference for students' development of their

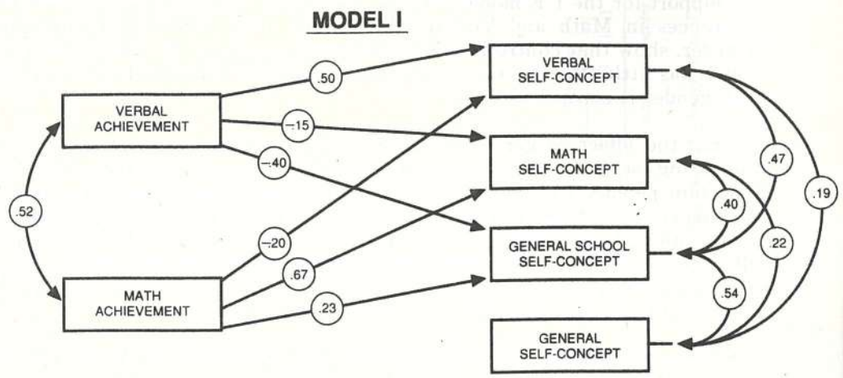


Figure 16A. Basic Internal–External Frame of Reference Model

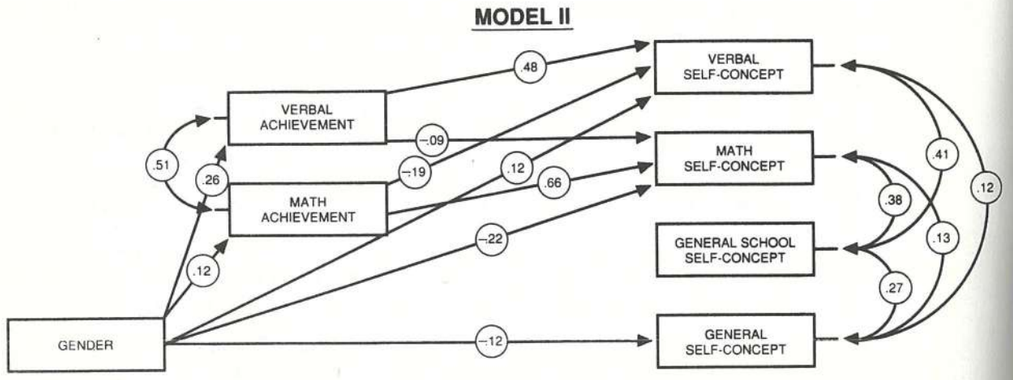


Figure 16B. Internal–External Frame of Reference Model with Gender

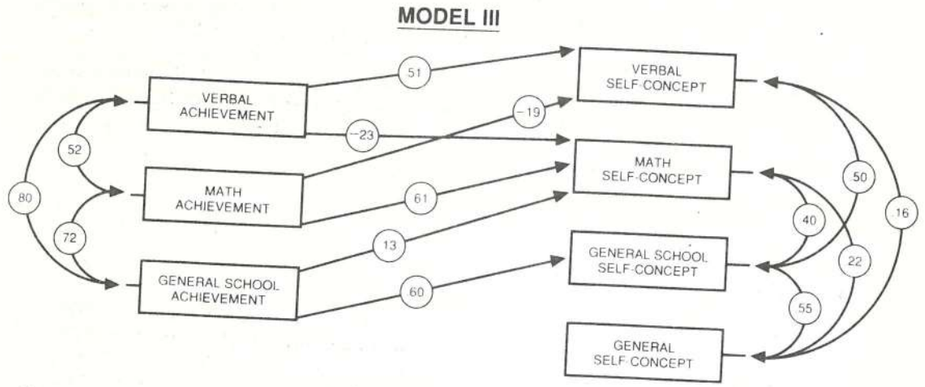


Figure 16C. Internal–External Frame of Reference Model with School Achievement

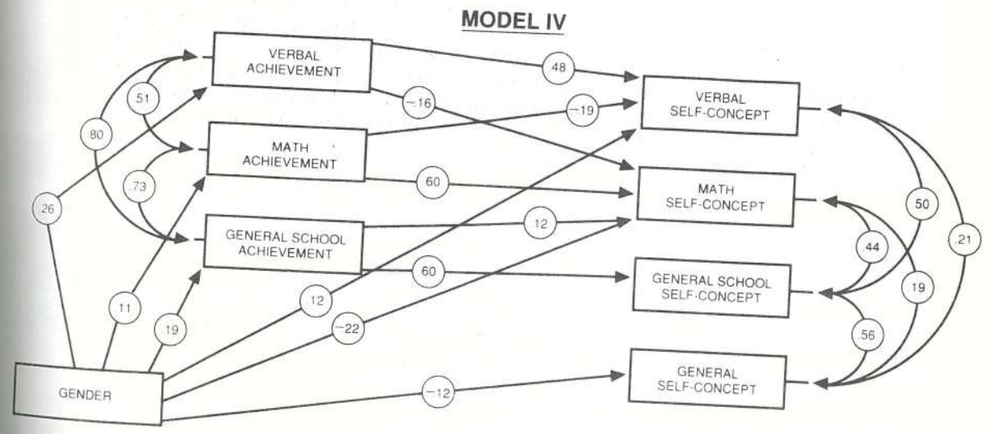


Figure 16D. Internal–External Frame of Reference Model with Gender and School Achievement

own self-concepts. Second, inferred self-concept ratings by teachers (and also by peers) overemphasize the external comparison of student skills in academic areas and underemphasize differences in skills in particular academic areas. Thus, a student who is weak in both math and verbal skills but who is stronger in one area than the other will tend to have much larger differences in Verbal and Math self-concepts than is reflected in the self-concepts inferred by teachers. The I-E model also predicts that nearly everyone will feel at least reasonably good about themselves in at least some areas.

Summary and Implications

Results summarized in this chapter demonstrate strong support for the construct validity of SDQ-II responses in relation to academic achievement indicators. Across a variety of different studies and different academic measures, academic self-concept scales were substantially related to academic criterion measures, whereas nonacademic self-concept scales were not. These findings demonstrate the importance of differentiating between academic and nonacademic facets of self-concept. Results summarized here also demonstrate, however, the need to distinguish between Math and Verbal self-concepts. Math and Verbal self-concepts were nearly uncorrelated with each other, and each had a distinctive pattern of relations to indicators of mathematics and verbal achievement. Overall measures of self-concept provide a poor basis of inference about academic self-concepts, and overall measures of academic self-concept typically provide a weak basis for inference about self-concepts in particular subject areas. These conclusions have important theoretical implications (e.g., the revision of the Shavelson et al. model) and practical implications for researchers (e.g., the inappropriateness of General self-concept for evaluating academic interventions) and for classroom teachers who want to better understand their students' self-concepts.

Chapter 8. Effects of Interventions

The enhancement of self-concept is widely viewed as a desirable goal and is frequently described as an intervening variable that may lead to desirable changes in other constructs (e.g., academic achievement). However, research reviewed earlier (see Chapter 5) indicates that self-concept is stable over time. Hence, it is not surprising that well-controlled interventions typically do not substantially affect self-concept, despite many possible biases that would be expected to produce changes in self-concept responses (e.g., placebo effects, acquiescence to the experimenter, postgroup euphoria).

Scheirer and Kraut (1979; also see Byrne, 1984) reviewed studies of academic interventions that attempted to improve self-concept as a means of improving academic achievement. They found predominantly nonsignificant results: most interventions failed to alter either self-concept or academic achievement. Similarly, Wylie (1979) reviewed studies of the effects of psychotherapy and growth-oriented group experiences on self-concept. She also found predominantly null results. Marsh, Richards, and Barnes (1986a, 1986b) suggested two possible reasons for this lack of success (also see Scheirer & Kraut, 1979; Wylie, 1979). First, most research is based on ill-defined measures of self-concept rather than on multidimensional measures in which some of the dimensions are specifically relevant to the aims of the intervention. Second, the size of the effect is typically small relative to probable error because the intervention is weak or because a potentially powerful intervention is administered only to a few subjects.

This chapter describes a construct validity approach to the study of intervention effects, as assessed by the SDQ-II. The construct validity of self-concept responses is typically tested in relation to (a) patterns of correlations with criterion variables (see, e.g., Marsh & Shavelson, 1985); (b) criterion group differences (Marsh & Jackson, 1986); and (c) the effects of experimental interventions designed to enhance self-concept (Marsh et al., 1986a, 1986b). Whereas intervention effects typically have not been used to test construct validity, the approach has been advocated for years by Cronbach (1970), Anastasi (1976), and many others.

There are important advantages to merging the methodologies of individual-difference research and experimental design. Such an approach fits nicely into the present emphasis on multidimensional self-concepts. Because interventions are typically relevant to only a few or even one dimension of self-concept, the intervention should have substantially more effect on the intervention-relevant dimensions than on the remaining dimensions. In this way, the remaining dimensions serve as a control for potential biases, such as placebo effects, that may invalidate construct interpretations. This construct validity approach is demonstrated in three SDQ studies. One of these interventions, the Outward Bound Standard Course, was evaluated with the SDQ-III and is described in the SDQ-III Manual (also see Marsh et al., 1986a, 1986b). A second intervention, the Outward Bound Bridging Course, was evaluated with the SDQ-I and is described in detail in

the SDQ-I Manual (also see Marsh & Richards, 1988a). These studies are summarized here only briefly. A third intervention, a physical fitness training program, was evaluated with the SDQ-II and is emphasized in this chapter (also see Marsh & Peart, 1988).

Study of the Outward Bound Standard Course

The Outward Bound Standard Course is a 26-day residential program for 17- to 25-year-olds which includes physically and mentally demanding outdoor activities. Richards (1977) stated that the purpose of the course is to provide a setting for "the person to recognize and understand his own weaknesses, strengths, and resources and thus find within himself the wherewithal to master the difficult and unfamiliar" (p. 69). Reviews of Outward Bound research (Godfrey, 1974; Richards, 1977; Shore, 1977) indicate that self-concept and locus of control are the most frequently examined outcome measures. Also, these reviews provide a theoretical basis for why the program should have an impact on these variables.

Emphasizing the multidimensionality of self-concept and a construct validity approach, Marsh et al. (1986b) argued that the specific dimensions of self-concept most logically related to this intervention should be most affected by it. They used a multiple-time-series design in which 27 different groups of participants completed the SDQ-III one month before the start of the program (Time 1, or T1), the first day of the program (T2), the last day of the program (T3), and 18 months after completing the program (T4). They found that (a) there was little change or a small drop in self-concept during the control (T1–T2) interval; (b) increases during the experimental (T2–T3) interval were significant and were significantly larger for those facets of self-concept judged a priori to be most relevant to the goals of the program; (c) increases during the experimental interval were similar for the 27 groups tested; and (d) there was little systematic shift in self-concepts during the follow-up (T3–T4) interval. Although responses to the *Rotter Internal–External Locus of Control Scale* (Rotter, 1966) were collected only at T2, T3, and T4, these responses showed locus of control to be more internal after completion of the program and also to remain stable during the follow-up interval.

Marsh et al. (1986b) argued that their design and the many checks on the construct validity of their interpretations provided a stronger test of the program effects than do most experimental designs. Their design included both a control interval and control facets of self-concept that were predicted to be less affected by the intervention. In particular, they argued that a randomly assigned control group would show whether or not self-concept is relatively stable over short periods of time and would provide little protection against potential biases such as placebo effects, Hawthorne effects, acquiescence to the experimenter, or postgroup-euphoria effects. Subsequent findings that the enhancement of self-concept was stable over the 18-month follow-up period support their contentions.

Study of the Outward Bound Bridging Course

The Bridging Course was developed by Outward Bound for low-achieving high school boys. Richards and Richards (1981) indicated that "the aim of the Outward Bound Bridging Course stated in its simplest form was to attempt to produce significant gains in the cognitive domain, especially in language and mathematics, through an integrated programme of remedial teaching, normal schoolwork and

experiences likely to influence personality in general and self-concept/self-esteem in particular" (p. 4). The design of the Bridging Course was influenced substantially by McClelland's (1965) achievement motivation theory and his practical suggestions about how achievement-motivation can be changed. The selection process is designed to create parental involvement and to engender a belief in the effectiveness of the program. The Bridging Course is a 6-week residential program in which a small group of 11 to 16 participants (primarily ninth-grade students) retreat to an isolated environment. The learning environment is task oriented, and teacher involvement with the students is emphasized. Educational materials are chosen to match the achievement levels of the participants. Initially, materials are below the achievement level of all participants, but they become progressively more difficult. The needs of individual students are determined, goals and criteria are clearly articulated, and individual student progress and performance are continuously assessed. Students are actively involved in the process of setting goals and monitoring their attainment of these goals in order to foster a sense of self-responsibility. Some of the outdoor physical activities from the traditional Outward Bound Standard Course are also included in the Bridging Course, but the primary focus is on an integrated approach to academic growth.

The Bridging Course was intended to enhance reading and mathematics achievement as well as to enhance the corresponding areas of academic self-concept. Although the program's effects on nonacademic areas of self-concept were expected to be smaller and less predictable on an a priori basis, the program was also expected to have some effect on nonacademic self-concepts. In particular, the substantial commitment made by parents suggested that effects on the Parent Relations scale might be positive. Nevertheless, it was predicted that the intervention would have more affect on academic areas of self-concept than on nonacademic areas.

The study used a short, multiple-time series design in which the SDQ-I and other measures were administered about 6 weeks before the start of the course (T1), on the first day of the course (T2), and on the last day of the course (T3). Neither academic achievement nor self-concept was likely to change systematically in such a short period without any intervention. Consistent with predictions, changes in self-concept during the T2–T3 experimental interval were substantial, statistically significant, more positive than corresponding changes in the control interval, and greater for academic than for nonacademic facets of self-concept. Also consistent with predictions, corresponding changes in reading and mathematics achievement occurred during the experimental interval. These findings provide strong support for (a) the effectiveness of the Outward Bound Bridging Course and parental support as an academic intervention for low-achieving high school boys with respect to both academic achievement and academic self-concept, (b) the multidimensionality of self-concept, (c) the validity of responses to the SDQ-I in relation to academic achievement, and (d) the effectiveness of the SDQ-I as a measure that validly reflects the effect of this academic intervention.

The Bridging Course study is one of the few studies in which a systematic intervention designed to enhance both academic achievement and academic self-concept was successful (see Scheirer & Kraut, 1979). The critical features that led to its success apparently were (a) a particularly powerful intervention which was conducted outside the school environment so that old self-concepts and behavior patterns would not be reinforced; (b) instilling expectations before the start of the intervention that changes would occur; (c) strong parental support for the program and parental expectations that the program would be successful, and (d) use of a multidimensional self-concept scale that not only validly measured areas of academic self-concept specific to the intervention's goals, but also differentiated these from other areas of self-concept. In their literature review, Marsh and Richards (1988a) were able to find only one other study in which an intervention had significant effects on both academic self-concept and academic achievement for

adolescent students—the classic study conducted by Brookover (see Brookover & Erikson, 1975; Scheirer & Kraut, 1979, for summaries). The design of the Brookover study is different from the present study in that it contained randomly assigned control and placebo subjects. In addition, the intervention used was also quite different. However, the four intervention characteristics identified above were also present in the Brookover research.

The results of the Bridging Course study complement those in the Standard Course study (Marsh et al., 1986b). In both studies, intervention programs encompassed in two different Outward Bound courses were found to enhance those self-concepts that were most specific to the aims of the respective courses. The juxtaposition of the two studies is particularly important. The Standard Course study predicted more change in the nonacademic than in the academic areas of self-concept because these areas were more relevant to its aims. This prediction was supported. In contrast, because of the academic aims of the Bridging Course, more change was predicted in the academic than in the nonacademic areas of self-concept. This prediction was also supported. In the Standard Course the aims of the interventions were specifically nonacademic, and the changes in the SDQ-III academic scales served as a control against which to evaluate SDQ-III facets that were more relevant. In the Bridging Course the aims of the intervention were primarily academic, and the changes in the SDQ-I nonacademic scales served as a control against which to evaluate changes in the SDQ-I academic scales. The Bridging Course study also differed in that changes in academic achievement provided an objective basis for assessing intervention effects and for validating changes that took place in academic self-concepts. Taken together, the two studies provide stronger support for the specificity of the effects of each of the interventions than is possible in considering either one in isolation.

Effects of Competitive and Cooperative Physical Fitness Training Programs

The study by Marsh and Peart (1988) had two emphases, one pragmatic and one theoretical. The pragmatic concern was to examine the differential effects of two physical fitness programs—one with a competitive orientation and one with a cooperative orientation—on physical fitness and on different facets of self-concept for high school girls. The theoretical concern was to examine further the construct validity of SDQ-II responses and the effects of interventions designed to affect self-concept.

Starting at an early age, girls participate in sports in smaller numbers, with less intensity, and for a shorter time span than do boys. Associated with this low participation is poor physical fitness and poor physical self-concepts. This finding may be examined in terms of different socialization patterns in that young girls are less likely to have positive role models for engaging in sports and receive less support from family and peers for engaging in sports (Snyder, Kivlin, & Spreitzer, 1979). The low participation rate may also be due, in part, to the competitive orientation of most high school sports and physical education programs. For example, Gill (1986) noted that "interpersonal competition is the dominant sport form today, especially in educational settings, and low levels of competitiveness may well be a major psychological barrier to sport participation" (p. 234). This competitive structure may differentially favor participation by boys in sports while discouraging participation by girls.

A competitive goal structure exists when certain students can achieve their own goals only when others fail to achieve theirs (i.e., there are winners and losers), whereas a cooperative goal structure exists when students can achieve their own individual goals only by working in cooperation with others as others achieve their goals (Deutsch, 1949; Johnson, Maruyama, Johnson, Nelson, & Skon, 1981; Owens & Straton, 1980). In a cooperative learning situation with intergroup competition, members within groups work together cooperatively even though there is competition between different groups, as in competitive team sports.

Reviews and meta-analyses have found that cooperation and cooperation with intergroup competition are more effective than interpersonal competition and individualistic goal structures for producing learning and achievement in many academic subject areas (Johnson et al., 1981; Sharan, 1980; Slavin, 1980) and for producing more positive academic attitudes, self-concepts, and related effects (Johnson & Johnson, 1983; Sharan, 1980; Slavin, 1980). Slavin examined whether the benefits of cooperative learning varied with either the sex of the student or the initial achievement level of students but found no clear patterns of interactions. High school girls, particularly in grades 8 and 9, prefer cooperative modes of learning in academic areas (Owens & Straton, 1980). Hence, particularly for girls, a cooperative learning environment may be both more effective and preferred by participants. A number of studies have examined the relationship between girls' involvement in sports and their self-concepts, but definition and measurement problems complicate the interpretations of these studies. Self-concept studies typically find that girls' athletic involvement has either little relation to, or a small positive relation to, overall self-concept (Ibrahim & Morrison, 1976; Snyder & Kivlin, 1975; Trujillo, 1983; Vincent, 1976; Young, 1981). It is likely, however, that some specific areas of self-concept will be substantially related to athletic involvement whereas others will not. Although some researchers (e.g., Ho & Walker, 1982; Snyder & Spreitzer, 1976) have found that women athletes have higher self-concepts than do women nonathletes in areas such as athletic ability and body image, two other studies (Ibrahim & Morrison, 1976; Young, 1981) found little or no difference on the Physical Self scale of the *Tennessee Self-Concept Scale* (Fitts, 1965; Roid & Fitts, 1988). Using the SDQ-III, Marsh and Jackson (1986) compared responses by high school female athletes and nonathletes, and by young adult female athletes and nonathletes. The athletes had substantially higher physical self-concepts than did the nonathletes, whereas differences on other scales were much smaller and typically failed to reach statistical significance. The specificity of the relations between athletic involvement and the multiple dimensions of self-concept supported the construct validity of the self-concept responses.

On the basis of Shavelson et al.'s hierarchical model of self-concept, Marsh and Jackson (1986) predicted—but did not find—that women's athletic involvement would be positively correlated with Physical Appearance self-concept. In discussing this unsupported prediction, they suggested that if the Physical Appearance scale had focused on physical fitness and body development, their prediction might have been supported. In retrospect, however, they noted that "self-perceived physical attractiveness—particularly for women—may not be related to athletic prowess and self-concepts of Physical Ability" (p. 412).

Consistent with this observation, recent tests of the hierarchical structure of responses to the SDQ-III, although generally supportive of the Shavelson et al. model, suggest that Physical Abilities and Physical Appearance self-concepts may not combine to form a second-order physical factor (Marsh, 1987c). This distinction may also explain why some researchers (e.g., Ibrahim & Morrison, 1976; Young, 1981) found little or no relation between female athletic involvement and the Physical Self scale of the *Tennessee Self-Concept Scale* which incorporates self-perceptions of body, health, appearance, skills, and sexuality into a single scale

(also see Marsh & Richards, 1988b). Hence, these results may provide further evidence against incorporating such diverse physical attributes into a single physical self scale.

Marsh and Peart (1988) examined the effects of a competitive versus a cooperative physical fitness training program on fitness and on SDQ-II responses. Subjects were 137 girls attending eighth grade at a Catholic girls' high school in metropolitan Sydney, Australia. They varied in age from 11 to 14 years (median = 13) and came from predominantly middle-class backgrounds. Few of the girls reported participating regularly in any extracurricular sporting activity prior to the study. Before and after participation in the intervention, all girls completed a physical fitness test composed of performances on seven exercises (e.g., running, push-ups, sit-ups) and the SDQ-II. In addition to the outcome measures used to assess the intervention effects, mathematics achievement indicators were available (see discussion in Chapter 7).

After completing the pretest measures, subjects were stratified in terms of initial physical fitness and randomly assigned to one of the two fitness programs or to a control group. The control group participated in an unstructured game of volleyball during the time when other students were participating in the two training programs. Although this activity was not carefully monitored, observation suggested that the girls treated this control activity as a social session in which there was little emphasis on competitive or strenuous activity.

The physical fitness program in which the two experimental groups participated was an aerobics training program designed to improve physical fitness. Each class session lasted approximately 35 minutes, and during the 6-week intervention there was a total of 14 sessions. Each session resembled traditional aerobics classes involving a mixture of calisthenics and athletics that were executed in time with music. The two groups differed in terms of the amount of cooperation required to perform the exercises and the verbal cues provided by the instructor. (The two programs and descriptions of the exercises unique to each program are presented in greater detail by Peart, 1985.) In the cooperative program, exercises were selected that required the cooperation of two girls, such that the successful performance by one participant required the cooperation of the other participant. Exercises in the competitive program could be performed individually. Specific verbal cues were developed to enhance the cooperative or the competitive nature of the respective programs.

The actual classes were conducted by two physical education teachers from the school and an aerobics instructor who was not otherwise connected with the school. Prior to the start of the intervention, the three teachers worked with one of the researchers to develop the exercises, the presentation of the instructional materials, and the type of feedback appropriate for each of the conditions. The three teachers were rotated every second class in order to counterbalance the effects of individual teachers. In both programs, the teacher demonstrated the exercises and then moved among the students, emphasizing either the cooperative or competitive nature of the session.

The study addressed the following hypotheses:

- The pretest measure of physical fitness would correlate more highly with Physical Abilities self-concept than with other self-concept scales, and the intervention would have more effect on the Physical Abilities than on the other self-concept scales.
- Both programs would enhance physical fitness compared to the control group's program, but the cooperative program would have more effect than the competitive program.

- The cooperative program—because it would enhance fitness, would be less threatening to the girls than the competitive program, and would conform to girls' preferences for cooperative instead of competitive learning environment—would enhance Physical Abilities self-concept.
- The competitive program—because it would have less effect on fitness, would be more threatening to girls than the cooperative program, and would not conform to girls' preferences for cooperative instead of competitive learning environments—would have a less positive effect on Physical Abilities self-concept than that of the cooperative group.

In the analysis of the preliminary results, the coefficient alpha estimates of reliability and test-retest correlations were first computed for the outcome measures. All alphas for the SDQ-II scales were .87 or higher except for the Honesty-Trustworthiness scale, which had alphas of .77 and .76. Coefficient alphas for the physical fitness test were .80 and .83 for T1 and T2, respectively. Stability coefficients (test-retest) for all variables varied from .72 to .88 during the 7 weeks between the pretest and posttest, even though there was an experimental intervention during this period (see the earlier discussion of Table 4, p. 30).

Of particular relevance to the construct validity emphasis of the present investigation is the pattern of correlations among the physical fitness test, the 11 SDQ-II scales, and the mathematics achievement score discussed earlier in Chapter 7 (see Table 11, p. 62). Consistent with predictions based on Marsh and Jackson's (1986) results, physical fitness was substantially correlated with Physical Abilities self-concept (.45) but not significantly correlated with any of the other 10 areas of self-concept. Consistent with previous self-concept research, mathematics achievement was substantially correlated with Math and General School self-concept (.48 and .50, respectively), less correlated with Verbal self-concept (.27), and even less correlated with each of the remaining eight SDQ-II scores (*rs* fell between −.16 and .23). The juxtaposition of the two sets of correlations provides strong support for the convergent and divergent validity of the multidimensional self-concept responses. The results also demonstrate the inappropriateness of assessing the effects of physical fitness interventions in terms of general or nonphysical components of self-concept.

What about the effects of the experimental intervention on the 12 outcome measures? In preliminary analyses no statistically significant group differences were found on any of the 12 pretest measures (all *ps* > .20). For each of the 12 posttest outcome measures, a 3 (experimental groups) × 4 (physical fitness pretest quartiles) analysis of covariance (ANCOVA) was conducted in which the corresponding pretest outcome measure was used as a covariate. The physical fitness pretest quartile did not interact significantly with the experimental manipulation for any of the 12 outcome variables, and thus it is not discussed further. Group means used in this analysis are presented in Table 15. These scores are average of responses to items on each SDQ-II scale, not the summative total scores.

Table 15. Effects of Participation in Alternative Physical Fitness Training Programs on SDQ-II Responses and on Physical Fitness

	Mean Item Responses for 3 Groups[a]									Significance Tests[b]			
	Cooperative (1)			Competitive (2)			Control (3)						
Outcomes	Time 1	Time 2	Adjust	Time 1	Time 2	Adjust	Time 1	Time 2	Adjust	*F*	1 *vs.* 2	1 *vs.* 3	2 *vs.* 3
Physical Fitness													
Fitness	−0.11	5.27	5.35	0.06	5.66	5.69	.07	.26	.13	57.69**	n.s.	**	**
Self-Concept Scales													
Physical Abilities	4.26	4.40	4.38	4.15	3.90	3.95	4.26	4.20	4.16	5.92**	**	n.s.	n.s.
Physical Appearance	3.29	3.57	3.53	3.16	3.00	3.05	3.23	3.25	3.24	4.12*	**	n.s.	n.s.
Opposite-Sex Relations	3.97	4.19	4.13	3.91	3.87	3.82	3.67	3.73	3.88	2.20	—	—	—
Same-Sex Relations	4.80	4.86	4.88	4.81	4.87	4.85	4.80	4.65	4.63	2.16	—	—	—
Parent Relations	4.56	4.45	4.61	4.78	4.70	4.68	4.96	4.71	4.53	0.41	—	—	—
Honesty–Trustworthiness	4.53	4.80	4.87	4.69	4.75	4.70	4.68	4.76	4.72	2.00	—	—	—
Emotional Stability	4.12	4.27	4.27	3.90	3.93	3.95	3.15	4.19	4.17	1.38	—	—	—
Math	3.20	3.25	3.33	3.37	3.28	3.21	3.36	3.32	3.29	0.42	—	—	—
Verbal	4.20	4.29	4.20	3.99	3.94	4.01	4.05	4.10	4.12	0.77	—	—	—
General School	4.17	4.18	4.20	4.08	3.95	4.03	4.34	4.18	4.05	0.94	—	—	—
General Self	4.43	4.61	4.56	4.27	4.28	4.39	4.45	4.48	4.41	1.06	—	—	—

[a]Time 1 = pretest scores; Time 2 = posttest scores; Adjust = posttest scores adjusted for pretest differences.
[b]For each of the outcome measures, a 3 (groups) × 4 (pretest fitness level) analysis of covariance was conducted in which the pretest measure corresponding to the posttest outcome measure was used as a covariate. Because there were no statistically significant interactions between groups and levels of pretest fitness (all *ps* > .05), these interaction effects are not summarized in the table.

$^{*}p < .05$; $^{**}p < .01$.

From "Competitive and Cooperative Physical Fitness Training Programs for Girls: Effects on Physical Fitness and Multidimensional Self-Concepts," by H. W. Marsh & N. D. Peart, 1988, *Journal of Sport Psychology, 10*, p. 400. Copyright 1988 by Human Kinetics Publishers. Adapted by permission of the publisher.

The experimental intervention had a large and substantial effect on physical fitness posttest scores. The competitive and cooperative groups each had substantially higher scores than did the control group, but they did not differ significantly from each other. Inspection of the adjusted posttest scores indicates that there was little pretest–posttest change for the control group, but a substantial improvement for the experimental groups. Whereas the physical fitness training program substantially affected fitness, there were no differences between the cooperative and competitive groups.

The experimental intervention had a significant effect on only 2 of the 11 self-concept mean responses—Physical Abilities and, to a lesser extent, Physical Appearance. For both these scales, posttest means were significantly higher for the cooperative group than for the competitive group. The posttest means for the cooperative group were higher than the corresponding pretest and the control group means, whereas the posttest means for the competitive group were lower than the corresponding pretest and control group means. The posttest mean for the control group was approximately midway between those of the competitive and cooperative groups. Thus, as predicted, the cooperative group did score significantly higher than the competitive group on two specific measures of physical self-concept.

For a better evaluation of the multivariate group differences, the entire set of outcome variables was considered simultaneously in a discriminant function analysis. These summarized multivariate group differences are consistent with those obtained with the univariate analyses of each separate outcome. The key difference is that the Physical Appearance self-concept was statistically significant for the univariate analyses, but it did not contribute significantly beyond the contribution of Physical Abilities self-concept in the multivariate analysis. These results are summarized in Table 16.

For purposes of the discriminant function analyses, adjusted posttest mean responses (a residual gain score in which pretest responses were partialed out of posttest responses) were used. Because of the large number of outcome variables, a stepwise selection procedure was used to assess the significance of each outcome measure to group differences. Both forward and backward selection procedures (see SPSS, 1988, for further discussion) indicated that only two outcome measures—physical fitness and Physical Abilities self-concept—contributed significantly to group differences.

There were two statistically significant discriminant functions: the first almost purely a function of physical fitness and the second almost purely a function of the Physical Abilities self-concept. All pairwise differences between groups were statistically significant (all *ps* < .01). On the first discriminant function (physical fitness), the two experimental groups had substantially more positive scores than did the control group. On the second discriminant function (Physical Abilities self-concept), the cooperative group was most positive, the control group was close to zero, and the competitive group was most negative. These multivariate and univariate analyses resulted in similar conclusions. In the multivariate analyses, however, Physical Appearance self-concept did not contribute significantly beyond the contribution of Physical Abilities self-concept.

In summary, the results of this study have important implications for the study of self-concept and for physical fitness training programs for girls. In support of the construct validity of the multidimensional self-concept responses, physical fitness was substantially correlated with Physical Abilities self-concept but not with other areas of self-concept. Also in support of this construct validity, an intervention program designed to influence Physical Abilities self-concept had a substantial effect on Physical Abilities self-concept but not on other areas of self-concept.

Table 16. Intervention Effects: Discriminant Analysis of Multivariate Group Differences on the Set of Outcome Measures Using Backward Elimination for Nonsignificant Variables

Variable	Standardized Function Coefficient		Correlation with Function	
	Function 1	Function 2	Function 1	Function 2
Physical Fitness				
Fitness	.999	.023	.999	.041
Self-Concept Scales				
Physical Abilities	−.041	.999	−.023	.999
Physical Appearance	—	—	—	—
Opposite-Sex Relations	—	—	—	—
Same-Sex Relations	—	—	—	—
Parent Relations	—	—	—	—
Honesty–Trustworthiness	—	—	—	—
Emotional Stability	—	—	—	—
Math	—	—	—	—
Verbal	—	—	—	—
General School	—	—	—	—
General Self	—	—	—	—
Group Means on Canonical Discriminant Functions				
Group 1	.469	.350		
Group 2	.614	−.346		
Group 3	−1.362	−.034		

Note: Physical fitness and self-concept of physical ability each contributed significantly to the discrimination between groups ($p < .01$), whereas the contribution of no other was statistically significant (all $ps > .05$). The final set of variables was the same for both forward and backward selection procedures. In the final solution, both discriminant functions were statistically significant, and pair-wise comparisons between the groups were all statistically significant ($p < .01$).

From "Competitive and Cooperative Physical Fitness Training Programs for Girls: Effects on Physical Fitness and Multidimensional Self-Concepts," by H. W. Marsh & N. D. Peart, 1988, *Journal of Sport Psychology, 10*, p. 402. Copyright 1988 by Human Kinetics Publishers. Adapted by permission of the publisher.

Cooperatively oriented physical fitness programs, as predicted, appear to be more beneficial for high school girls than are competitively oriented programs. Whereas both programs improved physical fitness, the cooperative program had a much more positive effect on Physical Abilities self-concept. The superiority of the cooperative program for enhancing Physical Abilities is not surprising; previous research has shown that girls prefer and are more oriented toward cooperative structures in general and particularly in sports. What may seem surprising is that Physical Abilities self-concept in the competitive program actually declined in comparison to pretest scores and in comparison to the control group. A possible explanation for these results is that the competitive nature of the program forced participants to compare their own physical abilities with those of the physically most able participants to a much greater degree than had been the case prior to the intervention or that occurred in the control group. In a setting where there are few winners and many losers, the average level of self-concept is likely to decline. Whereas participants were probably aware that they were more physically fit at the end of the competitive program, there was an even greater shift in the standard of comparison that they used to evaluate themselves.

In a related finding, Marsh and Parker (1984; Marsh, 1987a) described the "Big-Fish–Little-Pond effect": Students with average academic abilities have below-average academic self-concepts when school-average ability is high, and above-average academic self-concepts when school-average ability is low. In each case, self-concept is posited to reflect actual accomplishments and the frame of reference used to evaluate those accomplishments.

Summary and Implications

Well-controlled studies designed to enhance overall self-concept typically do not change self-concept. This is particularly true when global self-esteem is the only outcome variable. In many cases this may be due to inherent weakness in the actual intervention. If the intervention is ineffective, a lack of change in self-concept responses leads to a valid interpretation of the results. Even when the intervention is potentially powerful, failure to demonstrate its effect on self-concept responses may reflect problems with the self-concept instruments used to evaluate the interventions. Consistent with the construct validity approach emphasized in SDQ research, an intervention should have a substantially larger effect on specific dimensions of self-concept most relevant to the intervention and smaller or no effect on dimensions of self-concept that are less relevant. In this respect, the evaluations of interventions are substantially strengthened by using well-constructed multidimensional self-concept instruments such as the SDQ-II. Thus, the studies summarized in this chapter provide further support for the construct validity of the SDQ, including the scales designed to assess specific facets of self-concept.

Appendix A-1. Percentiles and *T* Scores for the Total Normative Sample

Raw Score	Physical Abilities %	Physical Abilities T	Physical Appearance %	Physical Appearance T	Opposite-Sex Relations %	Opposite-Sex Relations T	Same-Sex Relations %	Same-Sex Relations T	Parent Relations %	Parent Relations T	Honesty–Trust-worthiness %	Honesty–Trust-worthiness T	Emotional Stability %	Emotional Stability T	Math %	Math T	Verbal %	Verbal T	General School %	General School T	General Self %	General Self T	Raw Score
8			1	27																			**8**
9			2	28	≤1	≤23																	**9**
10			3	29	2	24			≤1	≤18					1	31							**10**
11			4	30	2	25			2	19					2	31							**11**
12	≤1	≤23	5	31	2	26			2	20					3	32							**12**
13	2	25	6	32	3	27			2	21					4	33							**13**
14	3	26	7	33	3	28			2	22					5	34							**14**
15	3	27	8	34	4	29			3	23	≤1	≤22	≤1	≤25	7	34	≤1	≤27					**15**
16	4	28	10	35	5	30			3	24	2	23	2	26	9	35	2	27	≤1	≤23			**16**
17	4	29	11	36	5	31			3	26	2	24	2	27	10	36	2	28	2	24			**17**
18	5	30	12	37	6	33			4	27	2	25	2	28	11	37	3	29	2	25			**18**
19	6	31	14	38	7	34			4	28	2	26	3	29	13	37	3	30	2	26	≤1	≤20	**19**
20	6	32	16	39	9	35			5	29	3	27	3	30	15	38	4	31	3	27	2	21	**20**
21	7	33	17	40	10	36			6	30	4	28	4	31	17	39	5	32	3	28	2	22	**21**
22	9	34	19	41	11	37			7	31	4	29	5	32	18	40	6	33	4	29	2	23	**22**
23	10	35	22	42	12	38			7	32	4	30	6	33	20	40	7	34	5	30	2	24	**23**
24	12	37	25	43	14	39			8	33	5	30	7	34	22	41	8	35	5	31	3	25	**24**
25	13	38	27	44	16	40			9	35	6	31	8	35	24	42	9	36	6	32	3	26	**25**
26	15	39	30	45	18	41	≤1	≤24	10	36	7	32	10	36	26	43	11	37	7	33	3	27	**26**
27	16	40	33	46	20	42	2	25	11	37	7	33	11	37	28	43	12	38	8	34	4	28	**27**
28	18	41	36	47	23	43	2	26	13	38	8	34	13	38	30	44	14	38	9	35	4	29	**28**
29	21	42	40	48	26	44	2	27	14	39	10	35	15	39	32	45	16	39	10	36	5	30	**29**
30	23	43	43	49	29	46	3	29	15	40	11	36	17	40	34	46	18	40	11	37	5	31	**30**
31	25	44	46	50	33	47	3	30	17	41	12	37	19	41	36	46	20	41	12	38	6	32	**31**
32	28	45	50	51	37	48	4	31	19	42	14	38	22	42	39	47	22	42	14	39	6	33	**32**
33	31	46	54	52	41	49	5	32	22	44	15	39	24	43	41	48	25	43	16	40	7	34	**33**
34	34	47	58	53	45	50	6	33	24	45	17	40	27	44	44	49	28	44	18	41	8	35	**34**
35	38	48	61	54	49	51	7	35	27	46	19	41	30	45	46	49	30	45	20	42	10	36	**35**
36	41	50	65	55	53	52	8	36	29	47	21	42	33	46	49	50	33	46	22	43	11	37	**36**
37	44	51	68	56	57	53	10	37	33	48	23	43	37	47	52	51	36	47	25	44	12	38	**37**

Note: $N = 5{,}494$.

Appendix A-1. Percentiles and *T* Scores for the Total Normative Sample *(continued)*

Raw Score	Physical Abilities %	Physical Abilities T	Physical Appearance %	Physical Appearance T	Opposite-Sex Relations %	Opposite-Sex Relations T	Same-Sex Relations %	Same-Sex Relations T	Parent Relations %	Parent Relations T	Honesty–Trust-worthiness %	Honesty–Trust-worthiness T	Emotional Stability %	Emotional Stability T	Math %	Math T	Verbal %	Verbal T	General School %	General School T	General Self %	General Self T	Raw Score
38	49	52	71	57	61	54	11	38	36	49	25	44	41	48	54	52	39	48	28	45	14	39	**38**
39	53	53	74	58	66	55	13	40	40	50	28	45	44	49	57	52	42	49	31	46	16	40	**39**
40	57	54	78	59	70	56	17	41	45	51	30	46	48	50	60	53	45	49	34	47	18	41	**40**
41	61	55	80	60	74	58	20	42	49	53	33	47	52	51	62	54	49	50	38	48	20	42	**41**
42	66	56	83	60	77	59	23	43	54	54	36	48	55	52	65	55	53	51	41	49	23	44	**42**
43	72	57	86	61	81	60	26	44	60	55	40	49	59	53	67	55	56	52	45	50	25	45	**43**
44	77	58	88	62	84	61	29	46	66	56	44	50	63	54	70	56	59	53	48	51	28	46	**44**
45	82	59	90	63	88	62	35	47	72	57	48	51	67	55	72	57	63	54	52	52	31	47	**45**
46	86	60	92	64	91	63	41	48	78	58	52	52	71	56	74	58	66	55	56	53	35	48	**46**
47	90	62	94	65	93	64	45	49	85	59	56	53	74	57	76	58	69	56	60	54	38	49	**47**
48	96	63	98	66	97	65	48	50	94	60	60	54	78	58	78	59	72	57	64	55	42	50	**48**
49							53	52			64	55	81	59	80	60	75	58	68	56	45	51	**49**
50							59	53			69	56	84	60	83	61	78	59	72	57	50	52	**50**
51							65	54			73	57	86	61	85	61	81	60	76	58	54	53	**51**
52							69	55			77	58	88	62	87	62	84	60	79	59	58	54	**52**
53							72	57			81	59	90	63	89	63	86	61	82	60	62	55	**53**
54							75	58			84	60	92	64	90	64	89	62	85	61	66	56	**54**
55							79	59			88	61	94	65	92	64	91	63	88	62	71	57	**55**
56							83	60			91	62	96	66	94	65	93	64	91	63	76	58	**56**
57							86	61			94	63	97	67	95	66	94	65	93	64	80	59	**57**
58							89	63			96	64	97	68	96	67	96	66	95	65	85	60	**58**
59							92	64			98	65	98	69	97	67	98	67	97	66	89	61	**59**
60							97	65			≥99	66	≥99	70	≥99	68	≥99	68	≥99	67	96	62	**60**
Mean	36.4		31.2		34.1		47.6		38.7		43.9		39.9		35.9		40.6		43.1		48.3		**Mean**
Median	38.0		32.0		35.0		48.0		41.0		46.0		40.3		36.0		41.0		44.0		50.0		**Median**
SD	9.2		10.3		9.2		8.2		8.9		10.2		10.1		13.3		10.9		10.2		9.7		***SD***
Skewness	−.85		−.33		−.59		−.70		−1.29		−.82		−.33		−.10		−.35		−.65		−1.13		**Skewness**
Kurtosis	.12		−.62		−.10		.85		1.33		.34		−.22		−.92		−.44		.14		1.41		**Kurtosis**

Note: $N = 5{,}494$.

Appendix A-2. Percentiles and *T* Scores: Males

Raw Score	Physical Abilities %	Physical Abilities *T*	Physical Appearance %	Physical Appearance *T*	Opposite-Sex Relations %	Opposite-Sex Relations *T*	Same-Sex Relations %	Same-Sex Relations *T*	Parent Relations %	Parent Relations *T*	Honesty–Trustworthiness %	Honesty–Trustworthiness *T*	Emotional Stability %	Emotional Stability *T*	Math %	Math *T*	Verbal %	Verbal *T*	General School %	General School *T*	General Self %	General Self *T*	Raw Score
8																							**8**
9					≤1	≤22																	**9**
10			≤1	≤25	2	23									≤1	≤29							**10**
11			2	26	2	24									2	30							**11**
12	≤1	≤20	2	27	2	25									3	30							**12**
13	2	21	3	28	3	26			≤1	≤19					3	31							**13**
14	2	22	4	29	3	27			2	20	≤1	≤25			4	32	≤1	≤27					**14**
15	2	23	4	30	4	28			2	21	2	26			6	33	2	28	≤1	≤23			**15**
16	3	24	5	31	4	29			2	22	3	27			7	33	2	29	2	24			**16**
17	3	26	6	32	5	30			3	23	3	27			8	34	3	30	2	25			**17**
18	3	27	7	33	6	32			3	25	3	28	≤1	≤26	9	35	3	31	2	26			**18**
19	4	28	8	34	7	33			3	26	4	29	2	27	10	36	4	32	3	27			**19**
20	4	29	9	35	8	34			4	27	4	30	2	28	12	37	5	33	3	28	≤1	≤20	**20**
21	5	30	10	36	9	35			4	28	5	31	3	29	13	37	6	34	4	29	2	21	**21**
22	6	31	12	37	10	36			5	29	6	32	4	30	15	38	7	35	4	30	2	22	**22**
23	7	33	14	38	11	37			6	31	6	33	4	31	16	39	8	35	5	31	2	23	**23**
24	8	34	16	39	12	38	≤1	≤24	7	32	7	34	5	32	17	40	9	36	6	32	2	24	**24**
25	9	35	18	40	14	39	2	25	8	33	9	35	6	33	19	40	11	37	7	33	3	25	**25**
26	11	36	20	41	16	40	2	26	8	34	10	36	7	34	21	41	13	38	8	34	3	26	**26**
27	12	37	22	43	18	41	2	28	9	35	11	37	8	35	23	42	15	39	8	34	4	27	**27**
28	14	38	25	44	20	42	3	29	11	37	12	37	9	36	25	43	17	40	9	35	4	28	**28**
29	16	40	28	45	23	43	3	30	13	38	14	38	11	37	27	43	19	41	10	36	4	29	**29**
30	17	41	31	46	26	44	4	31	14	39	16	39	12	38	30	44	22	42	12	37	5	30	**30**
31	20	42	35	47	30	45	4	32	16	40	18	40	14	39	32	45	24	43	13	38	5	31	**31**
32	22	43	38	48	33	47	5	34	18	41	20	41	16	40	34	46	27	44	15	39	6	32	**32**
33	25	44	42	49	37	48	7	35	20	43	22	42	18	41	37	47	30	44	16	40	6	33	**33**
34	28	45	46	50	41	49	8	36	22	44	24	43	20	42	39	47	32	45	18	41	7	34	**34**
35	30	47	50	51	44	50	10	37	25	45	26	44	23	43	41	48	36	46	20	42	9	35	**35**
36	33	48	53	52	48	51	12	38	28	46	28	45	26	44	44	49	39	47	23	43	10	36	**36**
37	37	49	57	53	52	52	14	39	31	47	31	46	30	45	47	50	42	48	25	44	11	37	**37**

Note: $N = 2{,}658$.

Appendix A-2. Percentiles and *T* Scores: Males *(continued)*

Raw Score	Physical Abilities %	Physical Abilities *T*	Physical Appearance %	Physical Appearance *T*	Opposite-Sex Relations %	Opposite-Sex Relations *T*	Same-Sex Relations %	Same-Sex Relations *T*	Parent Relations %	Parent Relations *T*	Honesty–Trustworthiness %	Honesty–Trustworthiness *T*	Emotional Stability %	Emotional Stability *T*	Math %	Math *T*	Verbal %	Verbal *T*	General School %	General School *T*	General Self %	General Self *T*	Raw Score
38	41	50	61	54	56	53	15	41	35	49	33	47	34	46	50	50	45	49	28	45	13	38	**38**
39	45	51	64	55	61	54	19	42	39	50	36	47	37	47	52	51	48	50	31	46	14	39	**39**
40	49	52	68	56	65	55	23	43	44	51	39	48	42	48	55	52	51	51	34	47	16	41	**40**
41	54	53	72	57	69	56	27	44	49	52	42	49	46	49	58	53	54	52	37	48	18	42	**41**
42	59	55	75	58	73	57	30	45	54	53	45	50	50	51	60	53	58	53	41	49	20	43	**42**
43	65	56	79	59	77	58	33	46	60	55	49	51	54	52	63	54	61	53	44	50	22	44	**43**
44	71	57	83	60	81	59	38	48	67	56	52	52	58	53	66	55	64	54	47	51	25	45	**44**
45	77	58	86	61	84	61	44	49	73	57	56	53	62	54	68	56	67	55	51	52	28	46	**45**
46	82	59	88	62	88	62	50	50	79	58	60	54	66	55	71	57	70	56	54	53	31	47	**46**
47	88	60	91	63	90	63	54	51	85	60	63	55	70	56	73	57	72	57	58	54	34	48	**47**
48	95	62	96	64	96	64	57	52	94	61	66	56	74	57	76	58	75	58	62	54	38	49	**48**
49							61	54			70	57	77	58	78	59	78	59	67	55	41	50	**49**
50							66	55			74	57	80	59	80	60	81	60	70	56	45	51	**50**
51							72	56			78	58	84	60	82	60	83	61	74	57	49	52	**51**
52							75	57			81	59	86	61	85	61	86	62	78	58	53	53	**52**
53							78	58			84	60	89	62	87	62	88	62	81	59	57	54	**53**
54							80	59			88	61	90	63	89	63	90	63	84	60	62	55	**54**
55							84	61			91	62	93	64	91	63	92	64	87	61	67	56	**55**
56							87	62			93	63	94	65	93	64	94	65	90	62	72	57	**56**
57							89	63			95	64	96	66	94	65	96	66	93	63	77	58	**57**
58							91	64			96	65	97	67	95	66	97	67	95	64	82	59	**58**
59							94	65			98	66	98	68	97	67	98	68	97	65	87	60	**59**
60							98	66			≥99	67	≥99	69	≥99	67	≥99	69	≥99	66	95	61	**60**
Mean	38.0		34.2		35.2		46.0		39.1		41.8		41.5		37.5		39.2		43.3		49.1		**Mean**
Median	40.0		35.0		36.0		46.0		41.0		43.0		42.0		38.0		40.0		45.0		51.0		**Median**
SD	8.6		9.6		9.3		8.5		8.3		11.0		9.6		13.0		11.1		10.5		9.6		***SD***
Skewness	−1.06		−.54		−.72		−.49		−1.31		−.61		−.40		−.20		−.23		−.67		−1.25		**Skewness**
Kurtosis	.74		−.32		.09		.32		1.65		−.15		−.02		−.82		−.60		.09		1.74		**Kurtosis**

Note: $N = 2{,}658$.

Appendix A-3. Percentiles and *T* Scores: Females

	Scale																						
	Physical Abilities		Physical Appearance		Opposite-Sex Relations		Same-Sex Relations		Parent Relations		Honesty–Trust-worthiness		Emotional Stability		Math		Verbal		General School		General Self		
Raw Score	%	*T*	%	*T*	%	*T*	%	*T*	%	*T*	%	*T*	%	*T*	%	*T*	%	*T*	%	*T*	%	*T*	Raw Score
8			2	30																			**8**
9			4	31					≤1	≤18													**9**
10			5	32	≤1	≤24			2	20					1	32							**10**
11	≤1	≤25	6	33	2	25			2	21					3	32							**11**
12	2	26	8	34	2	26			2	22					4	33							**12**
13	2	27	9	35	3	27			2	23					5	34							**13**
14	3	28	11	36	3	29			3	24			≤1	≤26	7	35							**14**
15	4	29	13	37	4	30			3	25			2	27	8	35							**15**
16	5	30	14	38	5	31			4	26			2	28	10	36	≤1	≤25					**16**
17	6	31	15	39	6	32			4	27			3	29	12	37	2	26	≤1	≤24			**17**
18	7	32	17	40	7	33			5	28			3	30	14	38	2	27	2	25	≤1	≤20	**18**
19	7	33	19	41	8	34			5	29			4	31	15	38	3	28	2	26	2	21	**19**
20	8	34	21	42	9	35			6	30	≤1	≤21	4	32	18	39	3	29	3	27	2	22	**20**
21	10	35	24	43	10	36			7	31	2	22	6	33	20	40	4	30	3	28	2	23	**21**
22	11	36	27	44	12	37			8	32	2	23	7	34	22	41	5	31	4	29	2	24	**22**
23	13	37	30	45	14	39			9	34	2	25	8	35	24	41	5	32	4	30	3	25	**23**
24	15	38	33	46	15	40			10	35	3	26	9	36	26	42	6	33	5	31	3	26	**24**
25	16	39	36	47	17	41			11	36	3	27	11	37	28	43	7	34	5	32	3	27	**25**
26	18	41	39	48	20	42			12	37	4	28	12	38	30	44	8	35	6	33	4	28	**26**
27	21	42	43	49	22	43			13	38	4	29	14	39	32	44	9	36	7	34	4	29	**27**
28	23	43	46	50	25	44			14	39	5	30	16	40	34	45	11	37	8	35	4	30	**28**
29	25	44	50	51	29	45	≤1	≤24	16	40	6	31	18	41	36	46	12	38	9	36	5	31	**29**
30	28	45	54	52	33	46	2	25	17	41	7	32	21	42	38	47	14	39	10	37	6	32	**30**
31	31	46	58	53	37	47	2	26	19	42	7	33	24	43	41	47	16	40	12	38	6	33	**31**
32	34	47	61	54	40	49	2	28	21	43	8	35	27	44	43	48	18	41	13	39	7	34	**32**
33	37	48	65	55	45	50	3	29	22	44	9	36	30	45	46	49	21	42	15	40	8	35	**33**
34	41	49	69	55	50	51	3	30	24	45	10	37	33	46	48	50	23	43	17	41	9	36	**34**
35	44	50	72	56	54	52	4	31	27	46	12	38	36	47	51	50	25	44	19	42	11	37	**35**
36	48	51	76	57	58	53	5	33	29	48	13	39	40	48	54	51	28	44	21	43	12	38	**36**
37	51	52	79	58	62	54	6	34	33	49	15	40	43	49	56	52	31	45	24	44	14	39	**37**

Note: $N = 2{,}836$.

Appendix A-3. Percentiles and *T* Scores: Females *(continued)*

	Scale																						
	Physical Abilities		Physical Appearance		Opposite–Sex Relations		Same-Sex Relations		Parent Relations		Honesty–Trust-worthiness		Emotional Stability		Math		Verbal		General School		General Self		
Raw Score	%	*T*	%	*T*	%	*T*	%	*T*	%	*T*	%	*T*	%	*T*	%	*T*	%	*T*	%	*T*	%	*T*	Raw Score
38	56	53	81	59	66	55	7	35	36	50	18	41	47	50	59	53	34	46	27	45	15	40	**38**
39	60	54	84	60	70	56	8	37	40	51	20	42	50	50	62	53	37	47	30	46	18	41	**39**
40	64	55	87	61	75	57	11	38	45	52	22	43	54	51	64	54	40	48	34	47	20	42	**40**
41	68	56	89	62	79	59	13	39	49	53	25	45	57	52	67	55	44	49	38	48	22	43	**41**
42	73	58	91	63	81	60	16	41	54	54	28	46	61	53	69	56	48	50	42	49	25	44	**42**
43	78	59	92	64	85	61	18	42	60	55	33	47	64	54	71	56	52	51	45	50	28	45	**43**
44	82	60	94	65	88	62	22	43	66	56	36	48	67	55	73	57	55	52	49	51	31	46	**44**
45	86	61	95	66	91	63	27	45	72	57	40	49	70	56	75	58	59	53	53	52	35	48	**45**
46	90	62	96	67	94	64	33	46	78	58	44	50	75	57	77	59	62	54	57	53	38	49	**46**
47	93	63	97	68	95	65	36	47	85	59	49	51	78	58	79	59	66	55	61	54	42	50	**47**
48	97	64	99	69	98	66	40	49	94	60	53	52	81	59	81	60	69	56	65	55	46	51	**48**
49							45	50			58	53	84	60	83	61	73	57	69	56	50	52	**49**
50							52	51			63	55	87	61	85	62	76	58	73	57	54	53	**50**
51							59	53			68	56	89	62	86	62	79	59	77	58	58	54	**51**
52							63	54			72	57	90	63	88	63	82	60	81	59	62	55	**52**
53							67	55			77	58	92	64	90	64	85	61	84	60	66	56	**53**
54							71	56			81	59	94	65	92	65	87	62	87	61	71	57	**54**
55							75	58			86	60	95	66	93	65	90	63	90	62	75	58	**55**
56							80	59			90	61	96	67	95	66	92	64	92	63	79	59	**56**
57							83	60			93	62	97	68	96	67	94	64	94	64	84	60	**57**
58							86	62			95	63	98	69	97	68	96	65	96	65	88	61	**58**
59							91	63			98	65	99	70	98	68	98	66	97	66	92	62	**59**
60							97	64			≥99	66	≥99	71	≥99	69	≥99	67	≥99	67	97	63	**60**
Mean	34.9		28.4		33.3		49.1		38.3		45.9		38.5		34.5		41.8		43.1		47.4		**Mean**
Median	37.0		29.0		34.0		50.0		41.0		47.0		39.0		35.0		43.0		44.0		49.0		**Median**
SD	9.4		10.2		9.0		7.6		9.3		9.0		10.4		13.4		10.5		9.9		9.7		***SD***
Skewness	−.68		−.15		−.53		−.93		−1.24		−.94		−.24		.01		−.45		−.65		−1.04		**Skewness**
Kurtosis	−.23		−.66		−.15		1.89		1.03		.94		−.36		−.96		−.24		.25		1.22		**Kurtosis**

Note: $N = 2{,}836$.

Appendix A-4. Percentile Equivalents of Total Self-Concept Raw Scores for the Total, Male, and Female Samples

Percentile	Total Sample	All Males	All Females	Percentile
≤1	126–268	146–273	116–262	≤1
2	269–291	274–294	263–288	2
3	292–307	295–309	289–304	3
4	308–321	310–320	305–315	4
5	322–328	321–330	316–329	5
6	329–334	331–338	330–334	6
7	335–342	339–343	335–338	7
8	343–346	344–349	339–343	8
9	347–351	350–354	344–347	9
10	352–355	355–358	348–352	10
11	356–358	359–363	353–356	11
12	359–363	364–366	357–359	12
13	364–366	367–369	360–362	13
14	367–369	370–373	363–365	14
15	370–372	374–377	366–368	15
16	373–376	378–381	369–372	16
17	377–379	382–383	373–374	17
18	380–383	384–385	375–377	18
19	384–386	386–388	378–381	19
20	387–388	389–391	382–384	20
21	389–390	392–393	385–387	21
22	391–392	394–395	388–391	22
23	393–394	396–397	392–392	23
24	395–396	398–400	393–393	24
25	397–398	401–402	394–395	25
26	399–401	403–404	396–398	26
27	402–403	405–406	399–400	27
28	404–405	407–408	401–402	28
29	406–407	409–410	403–403	29
30	408–409	411–413	404–405	30
31	410–411	414–415	406–406	31
32	412–413	416–416	407–408	32
33	414–414	417–418	409–410	33
34	415–415	419–420	411–413	34
35	416–416	421–422	414–414	35
36	417–418	423–424	415–416	36
37	419–419	425–426	417–417	37
38	420–421	427–428	418–418	38
39	422–425	429–431	419–420	39
40	426–427	432–432	421–423	40
41	428–428	434–434	424–424	41
42	429–430	435–435	425–426	42
43	431–431	436–437	427–427	43
44	432–433	438–440	428–428	44
45	434–435	441–441	429–430	45
46	436–436	442–443	431–431	46
47	437–438	444–445	432–433	47
48	439–440	446–447	434–435	48
49	441–441	448–450	436–436	49
50	442–443	451–452	437–438	50

Note: N = 5,494; all males = 2,658; all females = 2,836.

Appendix A-4. Percentile Equivalents of Total Self-Concept Raw Scores for the Total, Male, and Female Samples *(continued)*

Percentile	Total Sample	All Males	All Females	Percentile
51	444–445	453–453	439–439	51
52	446–447	454–455	440–441	52
53	448–448	456–456	442–443	53
54	449–450	457–458	444–444	54
55	451–452	459–459	445–446	55
56	453–454	460–461	447–447	56
57	455–456	462–463	448–449	57
58	457–458	464–465	450–451	58
59	459–460	466–466	452–453	59
60	461–462	467–468	454–454	60
61	463–463	469–470	455–455	61
62	464–465	471–472	456–457	62
63	466–466	473–473	458–460	63
64	467–468	474–475	461–461	64
65	469–470	476–476	462–463	65
66	471–471	477–478	464–465	66
67	472–473	479–479	466–466	67
68	474–476	480–481	467–468	68
69	477–477	482–482	469–470	69
70	478–479	483–484	471–472	70
71	480–480	485–486	473–474	71
72	481–482	487–488	475–476	72
73	483–484	489–490	477–478	73
74	485–486	491–492	479–481	74
75	487–487	493–495	482–484	75
76	488–490	496–497	485–486	76
77	491–492	498–498	487–487	77
78	493–495	499–501	488–489	78
79	496–497	502–503	490–490	79
80	498–500	504–505	491–492	80
81	501–501	506–507	493–495	81
82	502–502	508–509	496–497	82
83	503–504	510–511	498–499	83
84	505–507	512–514	500–501	84
85	508–509	515–516	502–504	85
86	510–512	517–518	505–507	86
87	513–515	519–519	508–509	87
88	516–517	520–523	510–513	88
89	518–520	524–526	514–517	89
90	521–523	527–529	518–520	90
91	524–526	530–532	521–522	91
92	527–530	533–536	523–526	92
93	531–535	537–541	527–531	93
94	536–541	542–547	532–537	94
95	542–547	548–550	538–542	95
96	548–552	551–555	543–547	96
97	553–558	556–562	548–554	97
98	559–567	563–568	555–564	98
≥99	568–612	569–612	565–612	≥99
Mean	439.8	444.8	435.2	**Mean**
Median	443.0	451.0	438.0	**Median**
SD	67.7	68.3	66.9	***SD***
Skewness	−0.48	−0.53	−0.45	**Skewness**
Kurtosis	0.52	0.40	0.68	**Kurtosis**

Note: N = 5,494; all males = 2,658; all females = 2,836.

Appendix A-5. *T*-Score Equivalents of Total Self-Concept Raw Scores for the Total, Male, and Female Samples

T Score	Total Sample	All Males	All Females	*T* Score
5	102–138	102–140	102–137	5
6	139–145	141–147	138–144	6
7	146–152	148–154	145–150	7
8	153–158	155–161	151–157	8
9	159–165	162–168	158–164	9
10	166–172	169–175	165–170	10
11	173–179	176–181	171–177	11
12	180–185	182–188	178–184	12
13	186–192	189–195	185–191	13
14	193–199	196–202	192–197	14
15	200–206	203–209	198–204	15
16	207–213	210–215	205–211	16
17	214–219	216–222	212–217	17
18	220–226	223–229	218–224	18
19	227–233	230–236	225–231	19
20	234–240	237–243	232–237	20
21	241–246	244–250	238–244	21
22	247–253	251–256	245–251	22
23	254–260	257–263	252–257	23
24	261–267	264–270	258–264	24
25	268–273	271–277	265–271	25
26	274–280	278–284	272–277	26
27	281–287	285–291	278–284	27
28	288–294	292–297	285–291	28
29	295–301	298–304	292–298	29
30	302–307	305–311	299–304	30
31	308–314	312–318	305–311	31
32	315–321	319–325	312–318	32
33	322–328	326–332	319–324	33
34	329–334	333–338	325–331	34
35	335–341	339–345	332–338	35
36	342–348	346–352	339–344	36
37	349–355	353–359	345–351	37
38	356–361	360–366	352–358	38
39	362–368	367–373	359–364	39
40	369–375	374–379	365–371	40
41	376–382	380–386	372–378	41
42	383–389	387–393	379–385	42
43	390–395	394–400	386–391	43
44	396–402	401–407	392–398	44
45	403–409	408–414	399–405	45
46	410–416	415–420	406–411	46
47	417–422	421–427	412–418	47
48	423–429	428–434	419–425	48
49	430–436	435–441	426–431	49

Note: $N = 5{,}494$; all males = 2,658; all females = 2,836.

Appendix A-5. *T*-Score Equivalents of Total Self-Concept Raw Scores for the Total, Male, and Female Samples *(continued)*

T Score	Total Sample	All Males	All Females	*T* Score
50	437–443	442–448	432–438	50
51	444–449	449–455	439–445	51
52	450–456	456–461	446–451	52
53	457–463	462–468	452–458	53
54	464–470	469–475	459–465	54
55	471–477	476–482	466–471	55
56	478–483	483–489	472–478	56
57	484–490	490–496	479–485	57
58	491–497	497–502	486–492	58
59	498–504	503–509	493–498	59
60	505–510	510–516	499–505	60
61	511–517	517–523	506–512	61
62	518–524	524–530	513–518	62
63	525–531	531–537	519–525	63
64	532–537	538–543	526–532	64
65	538–544	544–550	533–538	65
66	545–551	551–557	539–545	66
67	552–558	558–564	546–552	67
68	559–565	565–571	553–558	68
69	566–571	572–577	559–565	69
70	572–578	578–584	566–572	70
71	579–585	585–591	573–579	71
72	586–592	592–598	580–585	72
73	593–598	599–605	586–592	73
74	599–605	606–612	593–599	74
75	606–612		600–612	75

Note: $N = 5{,}494$; all males = 2,658; all females = 2,836.

Appendix B. Factor Score Coefficients (FSC) Based on the Total Normative Sample

Item Pair	Item Pair *M*	Item Pair *SD*	Scale Coefficients: Physical Abilities	Physical Appearance	Opposite-Sex Relations	Same-Sex Relations	Parent Relations	Honesty–Trustworthiness	Emotional Stability	Math	Verbal	General School	General Self
Physical Abilities 1	4.886	1.349	.352	-.032	-.015	.011	-.001	.021	-.016	-.002	.007	-.037	-.022
Physical Abilities 2	4.581	1.448	.282	-.013	.003	.002	-.005	-.027	-.007	-.008	.005	-.021	.009
Physical Abilities 3	4.315	1.366	.293	.003	-.018	-.020	-.003	.002	-.019	-.004	-.019	.009	-.026
Physical Abilities 4	4.427	1.284	.198	-.001	-.006	-.018	-.009	-.012	.008	-.006	-.005	.004	-.013
Physical Appearance 1	4.065	1.398	-.031	.260	.012	.043	-.007	.017	-.040	-.012	.002	-.008	-.005
Physical Appearance 2	3.599	1.470	-.009	.327	.000	-.046	-.003	-.021	.008	-.010	-.003	.019	-.066
Physical Appearance 3	3.923	1.429	-.017	.321	-.039	.010	.007	.016	-.004	.008	-.001	-.033	.009
Physical Appearance 4	4.029	1.476	.039	.224	-.032	-.052	-.013	-.003	-.023	-.005	-.007	-.014	.053
Opposite-Sex Relations 1	4.241	1.368	-.016	-.030	.267	.000	-.003	.016	-.010	.010	-.005	-.012	-.012
Opposite-Sex Relations 2	4.162	1.284	.008	.030	.269	-.030	.008	-.017	-.023	-.003	-.005	.017	-.060
Opposite-Sex Relations 3	4.581	1.255	.001	-.088	.307	-.008	-.002	.000	-.010	.005	.011	-.049	.029
Opposite-Sex Relations 4	4.080	1.363	-.019	.012	.290	-.006	-.011	.012	-.021	-.005	-.016	.013	-.030
Same-Sex Relations 1	4.751	1.044	.006	-.005	-.002	.326	-.002	-.020	-.009	.008	-.009	-.019	-.004
Same-Sex Relations 2	4.474	1.090	-.009	.032	-.005	.238	-.008	.005	-.010	-.004	-.002	.031	-.051
Same-Sex Relations 3	5.004	0.951	-.007	-.022	-.004	.254	-.001	-.003	-.004	.000	-.008	-.019	.014
Same-Sex Relations 4	4.799	1.020	-.010	-.012	.013	.219	-.005	.000	.017	.006	-.006	-.019	.000
Same-Sex Relations 5	4.785	1.019	-.001	-.015	-.017	.184	-.005	-.001	-.010	-.004	.001	.001	.011
Parent Relations 1	4.837	1.307	-.018	-.015	.009	.020	.377	.000	.022	-.009	-.007	-.006	-.015
Parent Relations 2	4.722	1.380	-.004	-.003	-.005	-.020	.305	-.045	.011	-.003	.002	-.028	-.012
Parent Relations 3	4.418	1.436	.001	.029	-.024	-.023	.249	.012	.035	.004	-.009	-.011	-.054
Parent Relations 4	5.367	1.087	-.002	-.025	.013	-.002	.193	-.011	-.062	-.011	-.009	-.002	.038
Honesty–Trustworthiness 1	4.786	1.237	-.008	-.013	.015	-.006	-.002	.297	.003	-.029	-.024	.010	.003
Honesty–Trustworthiness 2	3.539	1.388	-.010	.039	-.009	-.009	-.003	.219	.026	.004	.007	-.019	-.042
Honesty–Trustworthiness 3	4.585	1.417	.014	-.021	-.007	-.002	.008	.203	-.056	-.001	.005	-.006	.023
Honesty–Trustworthiness 4	4.499	1.281	.006	-.018	.004	.000	-.012	.265	-.010	-.002	.008	-.010	.002
Honesty–Trustworthiness 5	4.536	1.163	-.007	.016	.005	.009	-.008	.208	.006	-.011	-.012	-.004	.010
Emotional Stability 1	4.027	1.235	.001	-.009	-.002	-.004	-.010	.005	.218	-.006	-.001	-.005	.000
Emotional Stability 2	4.131	1.217	-.005	-.017	-.001	.044	.018	-.007	.203	.005	-.010	-.028	.012
Emotional Stability 3	4.174	1.241	.000	-.017	-.001	-.006	-.019	-.022	.227	-.007	.003	-.014	.019
Emotional Stability 4	3.837	1.468	-.002	.000	-.002	-.009	-.003	-.004	.308	.000	-.004	.002	-.026
Emotional Stability 5	3.803	1.354	-.005	.008	-.007	-.007	-.004	.029	.237	-.004	-.001	.005	-.028
Math 1	3.583	1.617	.002	-.002	.005	-.001	-.013	-.032	.044	.210	-.049	.073	-.046
Math 2	3.591	1.382	-.007	-.002	.004	.015	.004	.006	-.008	.213	.018	-.056	.002
Math 3	3.318	1.464	-.003	.014	-.003	-.012	.003	.015	-.003	.202	-.009	.008	-.019
Math 4	3.845	1.569	-.001	-.012	.003	-.006	.003	.002	-.023	.249	-.018	.049	.002
Math 5	3.626	1.574	.005	-.005	-.005	-.013	-.013	-.008	.020	.285	-.003	-.036	.016

Appendix B. Factor Score Coefficients (FSC) Based on the Total Normative Sample *(continued)*

Item Pair	Item Pair *M*	Item Pair *SD*	Scale Coefficients: Physical Abilities	Physical Appearance	Opposite-Sex Relations	Same-Sex Relations	Parent Relations	Honesty–Trustworthiness	Emotional Stability	Math	Verbal	General School	General Self
Verbal 1	3.873	1.445	.000	.087	-.006	-.024	.006	-.003	-.010	.015	.300	-.071	.004
Verbal 2	4.341	1.296	-.004	-.017	.014	.004	-.005	-.032	-.015	-.033	.199	.081	-.022
Verbal 3	4.060	1.215	-.006	.010	-.004	.007	-.009	-.001	-.040	-.002	.286	-.006	-.015
Verbal 4	4.137	1.396	-.002	-.015	-.009	.002	-.007	.012	.000	-.006	.229	.009	-.012
Verbal 5	3.910	1.395	-.001	-.002	.015	.004	-.011	-.013	.046	-.024	.165	.054	-.026
General School 1	3.812	1.380	-.004	.016	-.007	-.013	-.012	.007	.004	-.009	-.011	.178	-.006
General School 2	4.283	1.176	-.012	.025	-.021	.008	-.003	.004	-.043	-.006	.009	.140	.006
General School 3	4.571	1.272	.000	-.025	.006	-.019	-.001	.008	.001	-.015	.004	.189	.030
General School 4	4.309	1.195	.003	-.008	-.022	.007	-.004	-.036	-.023	-.017	.003	.306	-.011
General School 5	4.579	1.227	-.003	-.034	.001	-.011	-.003	-.024	.001	-.019	.018	.267	.029
General Self 1	4.833	1.199	-.016	.007	-.013	-.010	.000	-.001	-.015	-.011	-.023	-.021	.283
General Self 2	4.747	1.172	-.001	-.007	-.006	.000	-.013	-.013	.021	-.014	-.020	.033	.205
General Self 3	4.702	1.195	.012	-.017	-.008	-.019	-.014	-.005	.006	-.009	-.004	-.031	.294
General Self 4	4.819	1.177	.017	-.012	-.003	.018	.013	-.015	.022	-.014	-.011	-.001	.176
General Self 5	5.033	1.062	-.001	-.015	.000	-.002	-.001	-.016	-.027	-.015	-.017	.023	.217

Note: $N = 5{,}494$. The variables used to define the factor scores are the mean of responses to pairs of items; the first two items in each scale form the first item pair, the next two the second item pair, and so forth. In order to compute each factor score, use the equation: FACTOR SCORE = SUM [(X − MEAN)/*SD* × FSC)], where X = mean of responses to the items in a particular item pair for a given subject (see earlier discussion of item pairs), MEAN = item-pair mean given above, *SD* = item-pair standard deviation given above, and FSC = Factor Score Coefficients given above for each SDQ-II scale. Each factor score will consist of the sum of 51 products consisting of each item-pair *Z* score (i.e., (X − Mean)/*SD*) multiplied by its corresponding factor score. A total of 11 separate factor scores corresponding to the 11 SDQ scales are computed for each subject. Because the 11 factor scores are computed in relation to the normative sample, factor scores greater than 0.00 or less than 0.00 represent scores above or below the normative sample mean. Alternatively, the mean and *SD* from a user's own data can be used instead of the mean and *SD* based on the normative sample. In this case, the mean factor score should be approximately 0.00 for each factor.

References

Anastasi, A. (1976). *Psychological testing* (4th ed.). New York: Macmillan.

Antill, J. K., & Cunningham, J. D. (1979). Self-esteem as a function of masculinity in both sexes. *Journal of Consulting and Clinical Psychology, 47,* 783–785.

Antill, J. K., & Cunningham, J. D. (1980). The relationship of masculinity, femininity, and androgyny to self-esteem. *Australian Journal of Psychology, 32,* 195–207.

Antill, J. K., Cunningham, J. D., Russell, G., & Thompson, N. L. (1981). An Australian sex-role scale. *Australian Journal of Psychology, 33,* 169–183.

Bachman, J. G., & O'Malley, P. M. (1977). Self-esteem in young men: A longitudinal analysis of the impact of educational and occupational attainment. *Journal of Personality and Social Psychology, 35,* 365–380.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42,* 155–162.

Biggs, J. B. (1978). Individual and group differences in study processes. *British Journal of Educational Psychology, 48,* 266–279.

Biggs, J. B. (1984). *A survey of student's study processes.* Hawthorn, Victoria, Australia: Australian Council of Educational Research.

Boersma, F. J., & Chapman, J. W. (1979). *Student's Perception of Ability Scale manual.* Edmonton, Alberta, Canada: University of Alberta.

Brookover, W. B. (1962). *Self-Concept of Ability Scale.* East Lansing, MI: Educational Publication Services.

Brookover, W. B., & Erikson, E. L. (1975). *Sociology of education.* Homewood, IL: Dorsey Press.

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., & Shavelson, R. J. (1986). On the structure of adolescent self-concept. *Journal of Educational Psychology, 78,* 474–481.

Connell, W. F., Stroobant, R. E., Sinclair, K. E., Connell, R. W., & Rogers, K. W. (1975). *12 to 20: Studies of city youth.* Sydney, New South Wales, Australia: Hicks Smith & Sons.

Constantinople, A. (1973). Masculinity-femininity: An exception to a famous dictum? *Psychological Bulletin, 80,* 389–407.

Cook, E. P. (1985). *Psychological androgyny.* New York: Pergamon Press.

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

Cronbach, L. J. (1970). *Essentials of psychological testing.* New York: Harper & Row.

Deutsch, M. (1949). A theory of cooperation and competition. *Human Relations, 2,* 129–152.

Dusek, J. B., & Flaherty, J. F. (1981). The development of self-concept during adolescent years. *Monographs of the Society for Research in Child Development, 46* (4, Serial No. 191).

Eshel, Y., & Klein, Z. (1981). Development of academic self-concept of lower-class and middle-class primary school children. *Journal of Educational Psychology, 73,* 287–293.

Fitts, W. H. (1965). *Tennessee Self-Concept Scale: Manual.* Los Angeles: Western Psychological Services.

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.

Gill, D. L. (1986). Competitiveness among females and males in physical activity classes. *Sex Roles, 15,* 233–247.

Godfrey, R. (1974). *A review of research and evaluation literature on Outward Bound and related educational programs.* Denver, CO: Outward Bound School.

Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research, 52,* 123–142.

Harter, S. (1982). The Perceived Competence Scale for Children. *Child Development, 53,* 87–97.

Harter, S. (1983). Developmental perspectives on the self-system. In P. H. Mussen (Ed.), *Handbook of child psychology* (Vol. 4, 4th ed., pp. 275–385). New York: Wiley.

Hathaway, S. R., & McKinley, J. C. (1951). *The Minnesota Multiphasic Personality Inventory.* New York: The Psychological Corporation.

Ho, L., & Walker, J. F. (1982). Female athletes and nonathletes: Similarities and differences in self-perception. *Journal of Sport Behavior, 5,* 13–27.

Ibrahim, H., & Morrison, N. (1976). Self-actualization and self-concept among athletes. *Research Quarterly, 47,* 68–78.

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

Johnson, D. W., & Johnson, R. (1983). The socialization and achievement crisis: Are cooperative learning experiences the answer? *Applied Social Psychology Annual* (Vol. 4, pp. 5–54). Beverly Hills, CA: Sage.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, C., & Skon, L. (1981). The effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin, 89,* 47–62.

Kohlberg, L. (1966). A cognitive-developmental analysis of children's sex-role concepts and attitudes. In E. E. Maccoby (Ed.), *The development of sex differences* (pp. 82–173). Stanford, CA: Stanford University Press.

Lamke, L. K. (1982). The impact of sex-role orientation on self-esteem in early adolescence. *Child Development, 53,* 1530–1535.

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences.* Stanford, CA: Stanford University Press.

Marsh, H. W. (1985a). Age and sex effects in multiple dimensions of preadolescent self-concept. *Australian Journal of Psychology, 37,* 197–204.

Marsh, H. W. (1985b). The structure of masculinity/femininity: An application of confirmatory factor analysis to higher-order factor structures and factorial invariance. *Multivariate Behavioral Research, 20,* 427–449.

Marsh, H. W. (1986a). The bias of negatively worded items in rating scales for young children: A cognitive-developmental phenomenon. *Developmental Psychology, 22,* 37–49.

Marsh, H. W. (1986b). Global self-esteem: Its relation to weighted averages of specific facets of self-concept and their importance. *Journal of Personality and Social Psychology, 51,* 1224–1236.

Marsh, H. W. (1986c). The self-serving effect (bias?) in academic attributions: Its relation to academic achievement and self-concept. *Journal of Educational Psychology, 78,* 190–200.

Marsh, H. W. (1986d). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal, 23,* 129–149.

Marsh, H. W. (1987a). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology, 79,* 280–295.

Marsh, H. W. (1987b). The hierarchical structure of self-concept and the application of hierarchical confirmatory factor analysis. *Journal of Educational Measurement, 24,* 17–19.

Marsh, H. W. (1987c). Masculinity, femininity and androgyny: Their relations to multiple dimensions of self-concept. *Multivariate Behavioral Research, 22,* 91–118.

Marsh, H. W. (1988). *Self-Description Questionnaire-I: Manual and research monograph.* San Antonio, TX: The Psychological Corporation.

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

Marsh, H. W. (1989b). Sex differences in the development of verbal and mathematics constructs: The high school and beyond study. *American Educational Research Journal, 26,* 191–225.

Marsh, H. W. (in press). *The Self-Description Questionnaire-III: Manual and research monograph.* San Antonio, TX: The Psychological Corporation.

Marsh, H. W., Antill, J. K., & Cunningham, J. D. (1987). Masculinity, femininity, and androgyny: Relations to self esteem and social desirability. *Journal of Personality, 55,* 661–685.

Marsh, H. W., Barnes, J., Cairns, L., & Tidman, M. (1984). The Self-Description Questionnaire (SDQ): Age effects in the structure and level of self-concept for preadolescent children. *Journal of Educational Psychology, 76,* 940–956.

Marsh, H. W., Barnes, J., & Hocevar, D. (1985). Self-other agreement on multidimensional self-concept ratings: Factor analysis and multitrait-multimethod analysis. *Journal of Personality and Social Psychology, 49,* 1360-1377.

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., & Hocevar, D. (1985). The application of confirmatory factor analysis to the study of self-concept: First- and higher order factor structures and their invariance across age groups. *Psychological Bulletin, 97,* 562-582.

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., & Myers, M. R. (1986). Masculinity, femininity, and androgyny: A methodological and theoretical critique. *Sex Roles, 14,* 397-430.

Marsh, H. W., & O'Niell, R. (1984). Self-Description Questionnaire III: The construct validity of multidimensional self-concept ratings by late-adolescents. *Journal of Educational Measurement, 21,* 153-174.

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

Marsh, H. W., Parker, J., & Barnes, J. (1985). Multidimensional adolescent self-concepts: Their relationship to age, sex and academic measures. *American Educational Research Journal, 22,* 422-444.

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

Marsh, H. W., Relich, J. D., & Smith, I. D. (1983). Self-concept: The construct validity of interpretations based upon the SDQ. *Journal of Personality and Social Psychology, 45,* 173-187.

Marsh, H. W., & Richards, G. E. (1988a). The Outward Bound Bridging Course for low-achieving high school males: Effects 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 Scale: 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 followup of the effect of participation in an Outward Bound program. *Personality and Social Psychology Bulletin, 12,* 475-492.

Marsh, H. W., Richards, G. E., & Barnes, J. (1986b). Multidimensional self-concepts: The effect of participation in an Outward Bound program. *Journal of Personality and Social Psychology, 50,* 195-204.

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

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

Marsh, H. W., Smith, I. D., & Barnes, J. (1985). Multidimensional self-concepts: Relationships with sex and academic ability. *Journal of Educational Psychology, 77,* 581-596.

Marsh, H. W., Smith, I. D., Marsh, M. R., & Owens, L. (1988). The transition from single-sex to coeducational high schools: Effects on multiple dimensions of self-concept and on academic achievement. *American Educational Research Journal, 25,* 237-269.

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

McCarthy, J. D., & Hoge, D. R. (1982). Analysis of age effects in longitudinal studies of adolescent self-esteem. *Developmental Psychology, 18,* 372-379.

McClelland, D. C. (1965). Toward a theory of motive acquisition. *American Psychologist, 20,* 321-333.

Meece, J. L., Parsons, J. E., Kaczala, C. M., Goff, S. B., & Futterman, R. (1982). Sex differences in math achievement: Toward a model of academic choice. *Psychological Bulletin, 91,* 324-348.

Monge, R. H. (1973). Developmental trends in factors of adolescent self-concept. *Developmental Psychology, 8,* 382-393.

Montemayor, R., & Eisen, M. (1977). The development of self-conceptions from childhood to adolescence. *Developmental Psychology, 13,* 314-319.

Nicholls, J. (1979). Development of perception of attainment and causal attributions for success and failure in reading. *Journal of Educational Psychology, 71,* 94-99.

O'Malley, P. M., & Bachman, J. C. (1979). Self-esteem and education: Sex and cohort comparisons among high school seniors. *Journal of Personality and Social Psychology, 37,* 1153-1159.

O'Malley, P. M., & Bachman, J. C. (1983). Self-esteem: Change and stability between ages 13 and 23. *Developmental Psychology, 19,* 257-268.

Owens, L., & Straton, R. G. (1980). The development of a cooperative, competitive, and individualized learning preference scale for students. *British Journal of Educational Psychology, 50,* 112-126.

Peart, N. D. (1985). *Physical fitness for females: Dismantling the myth of "inferiority."* Unpublished bachelor's honors thesis, University of Sydney, Australia.

Pedhazur, E. J., & Tetenbaum, T. J. (1979). The Bem Sex-Role Inventory: A theoretical and methodological critique. *Journal of Personality and Social Psychology, 37,* 996-1016.

Piers, E. V. (1984). *Piers-Harris Children's Self-Concept Scale: Revised Manual 1984.* Los Angeles, CA: Western Psychological Services.

Piers, E. V., & Harris, D. A. (1964). Age and other correlates of self-concept in children. *Journal of Educational Psychology, 55,* 91-95.

Richards, G. E. (1977). *Some educational implications and contributions of Outward Bound.* Sydney: Australian Outward Bound Foundation.

Richards, G. E., & Richards, M. J. F. (1981). *Outward Bound Bridging Course 1981.* Sydney: Australian Outward Bound Foundation.

Roid, G. H., & Fitts, W. H. (1988). *Tennessee Self-Concept Scale-Revised Manual.* Los Angeles, CA: Western Psychological Services.

Rosenberg, M. (1965). *Society and the adolescent child.* Princeton, NJ: Princeton University Press.

Rosenberg, M. (1979). *Conceiving the self.* New York: Basic Books.

Rosenberg, M. (1985). Self-concept and psychological well-being in adolescence. In R. L. Leahy (Ed.), *The development of self* (pp. 55-121). Orlando, FL: Academic Press.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs, 80*(Whole No. 609).

Ruble, D. N., Boggiano, A. K., Feldman, N. S., & Loebl, J. H. (1980). Developmental analysis of the role of social comparison in self-evaluations. *Developmental Psychology, 16,* 105-115.

Scheirer, M. A., & Kraut, R. E. (1979). Increasing educational achievement via self-concept change. *Review of Educational Research, 49,* 131-150.

Sharan, S. (1980). Cooperative learning in teams: Recent methods and effects on achievement, attitudes, and ethnic relations. *Review of Educational Research, 50,* 241-272.

Shavelson, R. J., & Bolus, R. (1982). Self-concept: The interplay of theory and methods. *Journal of Educational Psychology, 74,* 3-17.

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

Shavelson, R. J., & Marsh, H. W. (1986). On the structure of self-concept. In R. Schwarzer (Ed.), *Anxiety and cognitions* (pp. 305-330). Hillsdale, NJ: Erlbaum.

Shore, A. (1977). *Outward Bound: A reference volume.* Greenwich, CT: Outward Bound, Inc.

Simmons, R. G., Rosenberg, F., & Rosenberg, M. (1973). Disturbance in the self-image at adolescence. *American Sociological Review, 38,* 553-568.

Slavin, R. E. (1980). Cooperative learning. *Review of Educational Research, 50,* 315-342.

Snyder, E. E., & Kivlin, J. E. (1975). Women athletes and aspects of psychological well-being and body image. *Research Quarterly, 46,* 191-199.

Snyder, E. E., Kivlin, J. E., & Spreitzer, E. E. (1979). The female athlete: An analysis of objective and subjective role conflict. In A. Yiannakis, T. D. McIntrye, M. J. Melnick, & D. P. Hart (Eds.), *Sport sociology: Contemporary themes* (pp. 204-216). Dubuque, IA: Kendall/Hunt.

Snyder, E. E., & Spreitzer, E. E. (1976). Correlates of sport participation among adolescent girls. *The Research Quarterly, 47,* 804-809.

Soares, A. T., & Soares, L. M. (1979). *The Affective Perception Inventory—Advanced Level.* Trumbull, CT: Soares Associates.

Soares, L. M., & Soares, A. T. (1977, April). *The self-concept: Mini, maxi, multi?* Paper presented at the 1977 Annual Meeting of the American Educational Research Association, New York.

Soares, L. M., & Soares, A. T. (1982, July). *Convergence and discrimination in academic self-concepts.* Paper presented at the 20th Congress of the International Association of Applied Psychology, Edinburgh, Scotland.

Spence, J. T. (1984). Masculinity, femininity, and gender-related traits: A conceptual analysis and critique of current research. In B. A. Maher & W. B. Maher (Eds.), *Progress in experimental personality research* (Vol. 13, pp. 1-97). New York: Academic Press.

Spence, J. T., Helmreich, R. L., & Holahan, C. K. (1979). Negative and positive components of masculinity and femininity and relations to self-reports of neurotic and acting out behaviors. *Journal of Personality and Social Psychology, 37,* 1673-1682.

Spielberger, C. D., Gonzalez, H. P., Taylor, C. J., Anton, W., Algaze, B., Ross, B., & Westberry, L. G. (1980). *Preliminary professional manual for the Test Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D., Gorsuch, R., & Luchene, R. (1970). *State-trait anxiety manual.* Palo Alto, CA: Consulting Psychologists Press.

SPSS, Inc. (1988). *SPSSx. User's guide.* Chicago: SPSS, Inc.

Stevenson, H. W., & Newman, R. S. (1986). Long-term prediction of achievement and attitudes in mathematics and reading. *Child Development, 57,* 646-659.

Stipek, D. J. (1981). Children's perceptions of their own and their classmates' ability. *Journal of Educational Psychology, 73,* 404-410.

Stipek, D. J. (1984). The development of achievement motivation. In R. E. Ames & C. Ames (Eds.), *Research on motivation in education* (Vol. 1, pp. 145-174). Orlando, FL: Academic Press.

Stipek, D. J., & Tannatt, L. M. (1984). Children's judgments of their own and their peers' academic competence. *Journal of Educational Psychology, 76,* 75-84.

Suls, J. M., & Miller, R. L. (1977). *Social comparison processes: Theoretical and empirical perspectives.* Washington, DC: Hemisphere Press.

Taylor, M. C., & Hall, J. A. (1982). Psychological androgyny: A review and reformulation of theories, methods, and conclusions. *Psychological Bulletin, 92,* 347-366.

Thomas, T. A. (1984). *An investigation into the influence of anxiety and the effectiveness of treatments involving positive coping skills to alleviate anxiety's detrimental effects.* Unpublished bachelor's honors thesis, University of Sydney, Australia.

Trowbridge, N. (1972). Self-concept and socio-economic status in elementary school children. *American Educational Research Journal, 9,* 525–537.

Trujillo, C. M. (1983). The effect of weight training and running exercise intervention programs on self-esteem of college women. *International Journal of Sport Psychology, 14,* 162–173.

Vincent, M. F. (1976). Comparison of self-concepts of college women: Athletes and physical education majors. *Research Quarterly, 47,* 218–225.

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

Werner, H. (1957). The concept of development from a comparative and organismic point of view. In D. B. Harris (Ed.), *The concept of development* (pp. 125–148). Minneapolis: University of Minnesota Press.

Whitley, B. E. (1983). Sex role orientation and self-esteem: A critical meta-analytic review. *Journal of Personality and Social Psychology, 44,* 765–778.

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

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

Young, M. (1981). Comparison of self-concepts of female high school and college tournament basketball players. *Research Quarterly, 36,* 490–495.